

Spatially inhomogeneous linear inverse problems with possible singularities

Marianna Pensky,

Department of Mathematics, University of Central Florida,
Orlando, FL 32816-1364, USA.

Email: Marianna.Pensky@ucf.edu

Abstract

The objective of the present paper is to introduce the concept of a spatially inhomogeneous linear inverse problem which, to the best of the author's knowledge, has never been considered previously in statistical framework but is emerging due to a variety of practical applications. The special feature of the problem is that the degree of ill-posedness depends not only on the scale but also on location. In this case, the rates of convergence are determined by the interaction of four parameters, the smoothness and spatial homogeneity of the unknown function f and degrees of ill-posedness and spatial inhomogeneity of operator Q . An interesting property here is that, if operator Q is weakly inhomogeneous, then the rates of convergence are not influenced by spatial inhomogeneity of operator Q and coincide with the rates which are usual for homogeneous linear inverse problems. On the other hand, if operator Q is moderately or strongly inhomogeneous, convergence rates are significantly affected by the degree of spatial inhomogeneity.

Estimators obtained in the paper are based either on wavelet-vaguelette decomposition (if the norms of all vaguelettes are finite) or on a hybrid of wavelet-vaguelette decomposition and Galerkin method (if vaguelettes in the neighborhood of the singularity point have infinite norms). The hybrid estimator is a combination of a linear part in the vicinity of the singularity point and the nonlinear block thresholding wavelet estimator elsewhere. To attain adaptivity, an optimal resolution level for the linear, singularity affected, portion of the estimator is obtained using Lepskii (1990, 1999) method. Subsequently, this resolution level is used as the lowest resolution level for the nonlinear wavelet estimator. We show that convergence rates of the hybrid estimator lie within a logarithmic factor of the optimal minimax convergence rates.

The theory presented in the paper is supplemented by examples of deconvolution with a spatially inhomogeneous kernel, deconvolution in the presence of locally extreme noise or extremely inhomogeneous design. The first two problems are examined via a limited simulation study which demonstrates advantages of the hybrid estimator when the degree of spatial inhomogeneity is high. In addition, we apply the technique to recovery of a convolution signal transmitted via amplitude modulation.

Keywords and phrases: Statistical linear inverse problems, inhomogeneous, minimax convergence rates, singularity

AMS (2000) Subject Classification: Primary: 62C10, 65J10. Secondary: 62G05, 62G20

1 Introduction

1.1 Formulation

Let Q be a known linear operator on a Hilbert space H with inner product $\langle \cdot, \cdot \rangle$. The objective is to recover $f \in H$ from observations on

$$y(x) = (Qf)(x) + \sqrt{\varepsilon}W(x), \quad x \in \mathcal{X}, \quad (1.1)$$

where $W(x)$ is the white noise process and $\sqrt{\varepsilon}$ is noise level. Assume that observations can be taken as functionals of y

$$\langle y, g \rangle = \langle Qf, g \rangle + \sqrt{\varepsilon} \xi(g), \quad g \in H, \quad (1.2)$$

where $\xi(g)$ is a Gaussian random variable with zero mean and variance $\|g\|^2$ such that $\mathbb{E}\xi(g_1)\xi(g_2) = \langle g_1, g_2 \rangle$. In what follows, $\|\cdot\|$ denotes the L^2 norm, all other norms are explicitly marked.

Model (1.1) is a common representation of a linear inverse problems with the Gaussian noise and has been studied by many authors (see, e.g., Abramovich and Silverman (1998), Bissantz *et al.* (2007), Cavalier and Golubev (2006), Cavalier *et al.* (2002), Cohen *et al.* (2004), Hoffmann, and Reis (2004), Donoho (1995), Golubev (2010), Hoffmann and Reiss (2008), Kalifa and Mallat (2003), and Mair and Ruymgaart (1996), among others). A typical assumption in the problem above is that operator Q acts uniformly over the spaces of functions represented at a common scale independently of the location of a function. In particular, consider a set of “test” functions

$$\psi_{ha}(x) = h^{-1/2} \psi\left(\frac{x-a}{h}\right) \quad (1.3)$$

where $\psi(x), x \in [0, 1]$, has a bounded support (L_ψ, U_ψ) and unit L^2 norm $\|\psi\| = 1$. Then, functions $\psi_{ha}(x)$ have scale h , supports concentrated around $x = a$ and unit norms. Conditions which are commonly imposed on operator Q imply that it contracts the norms of all functions ψ_{ha} uniformly, i.e., the value of $\|Q\psi_{ha}\|$ depends considerably on the scale h but hardly at all on a . Moreover, if there exist $(Q^*)^{-1}\psi_{ha}$, where Q^* is the adjoint of operator Q , then values of $\|(Q^*)^{-1}\psi_{ha}\|$ follow the same pattern. However, not all linear operators necessarily have those properties.

In order to illustrate the discussion above, consider linear operator Q with the adjoint Q^* given by

$$(Qf)(x) = \mu(x) \int_0^x f(t)dt, \quad (Q^*v)(x) = \int_x^1 \mu(z)v(z)dz, \quad (1.4)$$

where $\mu(x)$ is a smooth function. Assume that function ψ in (1.3) is continuously differentiable and integrates to zero: $\int_0^1 \psi(z)dz = 0$. Denote $\Psi(z) = \int_{L_\psi}^z \psi(x)dx$ and observe that $\Psi(z) = 0$ whenever $z \notin (L_\psi, U_\psi)$. Then, direct calculations yield

$$(Q\psi_{ha})(y) = h^{1/2} \mu(y) \Psi\left(\frac{y-a}{h}\right), \quad (Q^*)^{-1}\psi_{ha}(y) = -h^{-3/2} \mu^{-1}(y) \psi'\left(\frac{y-a}{h}\right).$$

so that

$$\begin{aligned} \|Q\psi_{ha}\|^2 &= h^2 \int_{L_\psi}^{U_\psi} \mu^2(a+hz) \Psi^2(z) dz = h^2 [\mu^2(a) \|\Psi\|^2 + o(1)], \quad (h \rightarrow 0), \\ \|(Q^*)^{-1}\psi_{ha}\|^2 &= h^{-2} \int_{L_\psi}^{U_\psi} \mu^{-2}(a+hz) [\psi'(z)]^2 dz. \end{aligned}$$

If $\mu(y)$ is a constant or, at least, $C_\mu^{-1} < \mu(y) < C_\mu$ for some relatively small C_μ , then dependence of $\|Q\psi_{ha}\|$ and $\|(Q^*)^{-1}\psi_{ha}\|$ on a can be ignored, and equation (1.1) with Q given by (1.4) can be treated as a spatially homogeneous problem. However, if $\mu(y)$ varies significantly, dependence on a becomes essential and equation (1.1) is a spatially inhomogeneous inverse problem.

Dependence on a becomes even more extreme if $\mu(y)$ vanishes at some point $x_0 \in (0, 1)$, e.g., $\mu^2(x) = C_\alpha |x - x_0|^\alpha$. Indeed, in this case, x_0 is the *singularity point* and it is easy to show that $\|Q\psi_{hx_0}\|^2 \asymp h^{2+\alpha}$ and

$$\|(Q^*)^{-1}\psi_{hx_0}\|^2 \asymp \begin{cases} h^{-(2+\alpha)}, & \text{if } \alpha < 1, \\ \infty, & \text{if } \alpha \geq 1. \end{cases}$$

Hence, if α is large, dependence of $\|Q\psi_{ha}\|$ and $\|(Q^*)^{-1}\psi_{ha}\|$ on location becomes quite extreme.

Since wavelets provide an adequate tool for scale-location representations of functional spaces, it is convenient to introduce spatially inhomogeneous linear inverse problems using a wavelet-vaguelette decomposition proposed by Donoho (1995). In particular, in the case when $H = L^2(\mathcal{D})$, $\mathcal{D} \subset R$, Donoho's assumptions appear as follows:

(D1) There exist three sets of functions: $\{\psi_{jk}\}$, an orthonormal wavelet basis of H , and nearly orthogonal sets $\{u_{jk}\}$ and $\{v_{jk}\}$ such that $Q\psi_{jk} = v_{jk}$, $Q^*u_{jk} = \psi_{jk}$, $\|v_{jk}\| = \lambda_j$, $\|u_{jk}\| \asymp \lambda_j^{-1}$, where λ_j depend on resolution index j but not on spatial index k .

(D2) u_{jk} and v_{jk} are such that $\langle u_{j_1k_1}, v_{j_2k_2} \rangle = \delta_{j_1,j_2} \delta_{k_1,k_2}$.

(D3) Sets $\{u_{jk}\}$ and $\{v_{jk}\}$ are nearly orthogonal, i.e., for any sequence $\{a_{jk}\} \in l^2$ one has

$$\left\| \sum_{j,k} a_{jk} \lambda_j u_{jk} \right\|^2 \asymp \sum_{j,k} a_{jk}^2, \quad \left\| \sum_{j,k} a_{jk} \lambda_j^{-1} v_{jk} \right\|^2 \asymp \sum_{j,k} a_{jk}^2.$$

Under conditions (D1)-(D3), f can be recovered using reproducing formula

$$f = \sum_{j,k} \langle Qf, u_{jk} \rangle \psi_{jk} \quad (1.5)$$

which is analogous to the reproducing formula for the SVD. Assumptions (D1)-(D3) are quite standard. Indeed, similar assumptions were introduced in Cavalier *et al.* (2002), Cavalier and Golubev (2006), Golubev (2010) and Knapik *et al.* (2011). The common premise is that operator Q acts "uniformly" over subspaces of H , so singular values or their surrogate equivalents depend on the resolution level only but not on location. If $V_j = \text{Span}\{\psi_{i,k}, i \leq j, k \in \mathbb{Z}\}$ is the subspace of functions at resolution level j , the above assumptions reduce to a common assumption of Galerkin method (see, e.g., Cohen *et al.* (2004) or Hoffmann and Reiss (2008)) that on subspace V_j operator Q has a bounded inverse with the norm dependent on j only, i.e., there exist $\lambda_j > 0$ such that

$$\sup_{j \geq 0} \left[\lambda_j \|Q^{-1}\|_{V_j \rightarrow Q^{-1}(V_j)} \right] < \infty, \quad (1.6)$$

which is very similar to combination of assumptions (D1) and (D3) above.

Note that both, assumptions (D1) and (1.6) imply that any function $v \in V_j$ with $\|v\| = 1$ has an inverse image, the norm of which is bounded by a constant which is independent of the support of v . In this sense, operator Q is an ill-posed *spatially homogeneous* operator. In the present paper, we shall be interested in a different situation when assumptions (D1) and (D3) may not be true. In particular, we assume that the norms of the inverse images of $\psi_{j,k}$ depend on the spatial index k and may be unbounded, i.e. condition (D1) and possibly condition (D3) are violated. We shall refer to the such inverse linear problems as *spatially inhomogeneous* in comparison with spatially homogeneous problems which satisfy conditions (D1)-(D3) above.

1.2 Motivation

Spatially inhomogeneous ill-posed problems appear naturally in the case when either the noise level is spatially dependent or observations are irregularly spaced. Problems of this kind have been considered before, both theoretically and in practical applications. Nevertheless, in former studies, it was always assumed that the noise level is uniformly bounded above or the design density of

observations is bounded away from zero. The situations investigated in the present paper rather refer to *locally extreme noise* and *extremely inhomogeneous design* (which can be also described as *a local data loss*). Traditionally, in the first situation, measurements are treated as outliers and are removed from future analysis while the second one is dealt with as the case of missing data. There are, however, multiple ill-posed problems where data quality varies and preserving all data for future analysis appears as a prudent choice. Problems of this sort deconvolution of LIDAR signals (see e.g., Harsdorf and Reuter (2000) and Gurdev *et al.* (2002)), or astronomical images (see, e.g., Starck *et al.* (2002) and Weddell and Webb (2008)) or analysis of forensic data (see, e.g., Li and Satta (2011)). Approach suggested in a present paper provides an alternative to missing data techniques which are usually applied in this case.

In addition, spatially inhomogeneous ill-posed problems arise in engineering or mathematical physics whenever the kernel is spatially inhomogeneous, as it occurs in the case of the amplitude modulation which is applied for transmitting information in the form of electro-magnetic waves.

Below we consider some examples in more detail.

Example 1 Deconvolution of LIDAR signals LIDAR (Light Detection And Ranging or Laser Imaging Detection And Ranging) is an optical remote sensing technology that can measure the distance to, or other properties of, targets by illuminating the target with laser light and analyzing the backscattered light. LIDAR technology has applications in archaeology, geography, geology, geomorphology, seismology, forestry, remote sensing, atmospheric physics. LIDAR data model is mathematically described by convolution equation $P = R * P_\delta$ where P is the time-resolved LIDAR signal, P_δ is the impulse response function and R is the system response function to be determined (see, e.g., Harsdorf and Reuter (2000) and Gurdev *et al.* (2002)). However, if the system response function of the LIDAR is longer than the time resolution interval, then the measured LIDAR signal is blurred and the effective accuracy of the LIDAR decreases. This loss of precision becomes extreme when, for example, LIDAR is used to for emergency response and natural disaster management such as assessment of the extent of damage due to volcanic eruptions or forest fires. This is due to the presence of dust, smoke and other obstructions which LIDAR signal cannot penetrate. In this situation, routinely, distances are calculated through filtering of the data set (removing outliers) and applying interpolation techniques. However, keeping all existing data and accounting for extreme noise may improve precision of the analysis of LIDAR signals.

Example 2 Deconvolution of astronomical data Deconvolution of astronomical images has proven in some cases to be crucial for extracting scientific content. For example, deconvolved mid-infrared images are used to reveal inner structure of the active galactic nucleus hidden at lower wavelength because of the high extinction. Also, research on gravitational lenses is easier and more efficient when applying deconvolution methods (see, e.g., Starck *et al.* (2002) and references therein). In addition, deconvolution is also crucial in order to fully take advantage of increasing numbers of high-quality ground-based telescopes like the Hubble Space Telescope, for which images are strongly limited in resolution by the seeing.

Analysis of astronomical images is usually formulated as a two-dimensional deconvolution problem with the spatial impulse response function, commonly referred to as the point spread function (PSF), as a kernel and an additive noise. Extreme measurement errors are ubiquitous in astronomy. Common sources of measurement error are the Poissonian nature of photon counts, instrumental noise, and calibration. In addition to the ever-present effect of noise from imaging equipment and optical defects from instrumentation, images from ground-based telescopes are distorted by wavefront aberrations caused by atmospheric turbulence. The PSF which is used to represent such distortions can either be applied over the entire image, or within regions uniquely defined by the isoplanatic angle. The combination of such regions forms an extended image, where the spatially variant PSF is used for image restoration (see, e.g., Weddell and Webb (2008)). Both

situations lead to a two-dimensional version of model considered in Section 7.1 where large degrees of spatial inhomogeneity correspond to extreme distortions.

Example 3 Amplitude Modulation Amplitude Modulation (AM) is a way of transmitting information in the form of electro-magnetic waves. In AM, a radio wave known as the "carrier" or "carrier wave" is modulated in amplitude by the signal that is to be transmitted, while the frequency remains constant (see, e.g., Miller *et al.* (2009)). In video or image transmission (such as TV) where the base-band signal has inherent large bandwidth, AM is usually preferred to Frequency Modulation (FM) systems since the latter ones require additional bandwidth. Since in an AM, signal information is "stored" in amplitude which is affected by noise, AM is more susceptible to noise than FM. Mathematically, the problem reduces to multiplying the transmitted signal by the function $\mu(x) = \cos(2\pi\omega x - \theta)$ with large $\omega \approx n/2$ and $\theta \in [0; 2\pi]$. In Section 8.2 we provide an in-depth description of application of the methodology developed in the paper to recovery of a convolution signal transmitted via AM.

1.3 Objectives and layout of the paper

The objective of the present paper is to introduce the concept of a spatially inhomogeneous linear inverse problem which, to the best of the author's knowledge, has never been considered previously in statistical framework but is emerging due to a variety of practical applications. It turns out that spatially inhomogeneous problems exhibit properties which are very different from their spatially homogeneous counterparts. In particular, if the norms of vaguelettes $u_{j,k} = (Q^*)^{-1}\psi_{jk}$ are infinite in the vicinity of a singularity point, reproducing formula (1.5) cease working and the usual wavelet-vaguelette estimators cannot be applied. In this case, we propose a hybrid estimator which is based on combination of wavelet-vaguelette decomposition and Galerkin method. We study two application of the general theory, deconvolution with spatially inhomogeneous design and deconvolution with a spatially inhomogeneous kernel (the case of heterogeneous noise being a particular case of the latter).

Another interesting feature of the model is that the rates of convergence are determined by the interaction of four parameters, the smoothness and spatial homogeneity of the unknown function f and degrees of ill-posedness and spatial inhomogeneity of operator Q . In particular, if operator Q is weakly inhomogeneous, then the rates of convergence are not influenced by spatial inhomogeneity of operator Q and coincide with the rates which are usual for homogeneous linear inverse problems.

In what follows, we assume that operator Q in (1.1) is completely known. If, in practical applications, this is not true, one has to account for the extraneous errors which stem from the uncertainty in the operator Q by using, for example, ideas of Hoffman and Reiss (2008). Also, to simplify our considerations, we limit our study to the case when $\mathcal{X} = [0, 1]$, $H = L^2[0, 1]$, and k is a scalar. The theory presented below can be generalized to the case when $H = L^2(\mathcal{D})$, $\mathcal{D} \subset R^d$ and k is a d -dimensional vector. This extension should be relatively straightforward if one is dealing with isotropic Besov spaces but becomes much more interesting and involved in the case of anisotropic Besov spaces (see, e.g. Kerkycharian, Lepski and Picard (2001)). However, we leave those extensions for future investigations since considering them below will prevent us from focusing on the main objective of the paper.

The rest of the paper is organized as follows. Section 2 introduces the concept of a spatially inhomogeneous ill-posed problem and formulates major definitions and assumptions which are used throughout the paper. Section 3 presents the asymptotic minimax lower bounds for the L^2 -risk of the estimators of the solution of the problem over a wide range of Besov balls. Section 4 talks about estimation strategies, in particular, about partitioning the unknown response function f and its estimator into the singularity-affected and the singularity-free parts, the main idea at the core of the hybrid estimator. Section 5 elaborates on the risk of the estimator constructed in the

previous section when the lowest resolution level in the zero-affected portion of the estimator is fixed. Section 6 discusses the adaptive choice of the lowest resolution level and derives the asymptotic minimax upper bounds for the L^2 -risk. In Section 7, we consider two examples of spatially inhomogeneous ill-posed problems, deconvolution with the spatially inhomogeneous operator (Section 7.1) which can be viewed as a version of a deconvolution equation with spatially inhomogeneous noise, and deconvolution based on irregularly spaced sample (Section 7.2). Section 8 presents a limited simulation study of deconvolution with heteroscedastic noise which demonstrates advantages of the hybrid estimator when the degree of spatial inhomogeneity is high. Section 8 also studies application of the hybrid estimator to recovery of a convolution signal transmitted via Amplitude Modulation. Section 9 concludes the paper with a discussion. Finally, Section 10 contains the proofs of the statements in the earlier sections.

2 Spatially inhomogeneous ill-posed problem: assumptions and definitions

Consider a scaling function φ and a corresponding wavelet ψ with bounded supports and form an orthonormal wavelet basis $\{\psi_{jk}\}$ of $L^2([0, 1])$. We further impose the following set of assumptions on spatially inhomogeneous operator Q .

(A1) There exist functions $\{u_{j,k}\}$ and $\{v_{j,k}\}$ such that

$$Q\psi_{jk} = v_{j,k}, \quad Q^*u_{j,k} = \psi_{jk}, \quad (2.1)$$

where $\|v_{j,k}\| = \lambda_{j,k} < \infty$.

(A2) There exists a *singularity* point $x_0 \in (0, 1)$ and a constant $D \geq 0$ such that $\|u_{j,k}\| = \infty$ if $|k - k_{0j}| < D$ and for any $\{a_{j,k}\}$, $k = 0, \dots, 2^j - 1$, one has

$$\left\| \sum_{|k - k_{0j}| \geq D} a_{j,k} \lambda_{j,k} u_{j,k} \right\|^2 \leq C_u \sum_{|k - k_{0j}| \geq D} a_{j,k}^2, \quad (2.2)$$

where $C_u < \infty$ is independent of j and $k_{0j} = 2^j x_0$ is the parameter corresponding to location x_0 (k_{0j} is not necessarily an integer).

(A3) Functions $v_{j,k}$ are such that inequality

$$\left\| \sum_{k=0}^{2^j-1} a_{j,k} \lambda_{j,k}^{-1} v_{j,k} \right\|^2 \leq C_v \sum_{k=0}^{2^j-1} a_{j,k}^2 \quad (2.3)$$

holds for any $\{a_{j,k}\}$, $k = 0, \dots, 2^j - 1$, where $C_v < \infty$ is independent of j .

Note that Assumptions (A1)–(A3) are much weaker than Assumptions (D1)–(D3) above. First, $\lambda_{j,k}$ depends not only on resolution level but also on location of the wavelet coefficient. Observe also that, even if $D = 0$, assumptions (A2) and (A3) are weaker than assumption (D3) since the sums is taken over one resolution level only. Moreover, if $D > 0$, then, in the neighborhood of the singularity point x_0 , wavelet coefficients cannot be recovered directly since $\|u_{j,k}\| = \infty$, and we say that operator Q has a *singularity* at x_0 .

Since one usually start wavelet expansion at some finite resolution level m , below we list an extra assumption which mirrors Assumption (A2) and can be derived from it:

(A4) There exist functions $\{t_{m,k}\}$ and positive constants C_t and D_0 independent of m such that for any $\{a_{mk}\}$, $k = 0, \dots, 2^m - 1$,

$$Q^* t_{m,k} = \varphi_{mk}, \quad \left\| \sum_{|k-k_{0j}| \geq D_0} a_{mk} \lambda_{m,k} t_{m,k} \right\|^2 \leq C_t \sum_{|k-k_{0j}| \geq D_0} a_{mk}^2. \quad (2.4)$$

If $D = D_0 = 0$ in Assumptions (A2) and (A4), then $\|u_{j,k}\| < \infty$ and $\|t_{m,k}\| < \infty$ for any k . Hence, f can be expressed using reproducing formula (1.5) which, in this case, becomes

$$f(x) = \sum_{k=0}^{2^m-1} a_{mk} \varphi_{mk}(x) + \sum_{j=m}^{\infty} \sum_{k=0}^{2^j-1} b_{jk} \psi_{jk}(x), \quad (2.5)$$

where $a_{mk} = \langle Qf, t_{m,k} \rangle$ and $b_{jk} = \langle Qf, u_{j,k} \rangle$. If $D > 0$, reproducing formula (2.5) does not work and one needs an alternative solution to recovering f . Indeed, if Qf in expressions for a_{mk} and b_{jk} is replaced by $y = Qf + \varepsilon W$, then the variances of the wavelet coefficients in the vicinity of singularity x_0 are infinite: $\text{Var}\langle y, u_{j,k} \rangle = \infty$ if $|k - k_{0j}| < D$ and similar consideration applies to $t_{m,k}$. For this reason, at each resolution level, we partition the set of all indices into the *singularity-affected* indices

$$K_{0m} = \{k = 0, \dots, 2^m - 1 : |k - k_{0m}| < D_0\}, \quad K_{1j} = \{k = 0, \dots, 2^j - 1 : |k - k_{0j}| < D\}$$

and the *singularity-free* indices

$$K_{0m}^c = \{k : 0 \leq k \leq 2^m - 1, k \notin K_{0m}\}, \quad K_{1j}^c = \{k : 0 \leq k \leq 2^j - 1, k \notin K_{1j}\}.$$

To be specific, in what follows, we assume that $\lambda_{j,k}$ are such that, for some positive constants α, β , C_{λ_0} and C_λ independent of j and k , one has

$$C_{\lambda_0} 2^{-j(\alpha+\beta)} (1 + |k - k_{0j}|)^\alpha \leq \lambda_{j,k}^2 \leq C_\lambda 2^{-j(\alpha+\beta)} (1 + |k - k_{0j}|)^\alpha. \quad (2.6)$$

We shall refer to coefficients β and α in (2.6) as degrees of *ill-posedness* and *spatial inhomogeneity*, respectively. Observe that with $\lambda_{j,k}$ satisfying condition (2.6), the variances of the coefficients at the lower resolution levels may be significantly higher than the variances of the coefficients at higher resolution levels as long as the location of the lower resolution level coefficients lie in a close proximity of a singularity point.

In the present paper, we consider estimation of a solution of inhomogeneous linear inverse problems in the case when the unknown function f is possibly spatially inhomogeneous itself, in particular, f belongs to a Besov ball $B_{p,q}^s(A)$ of radius A . Interplay between spatial inhomogeneity of operator Q and properties of f lead to various very interesting phenomena. In particular, if α is small or p and β are relatively large, spatial inhomogeneity does not affect convergence rates and f can be recovered as well as in the case of $\alpha = 0$.

Remark 1 (Multiple singularity points) Note that one can consider a spatially inhomogeneous problems with multiple singularity points $x_{0,1} < x_{0,2} < \dots < x_{0,L}$ and corresponding constants D_1, \dots, D_L where $L < \infty$ and $x_{0,i} - x_{0,i-1} \geq \delta > 0$ for some fixed positive δ . Theory developed below can be easily extended to this case, with the convergence rates of the estimators determined by the "worst case scenario" among singular points $x_{0,i}$, $i = 1, \dots, L$.

3 Minimax lower bounds for the risk over Besov balls

Before constructing an estimator of the unknown function f under model (1.1), we derive the asymptotic minimax lower bounds for the L^2 -risk over a wide range of Besov balls.

Recall that for an r_0 -regular multiresolution analysis (see, e.g., Meyer (1992), pp 21–25), with $0 < s < r_0$, and for a Besov ball $B_{p,q}^s(A)$

$$B_{p,q}^s(A) = \{f \in L^p([0, 1]) : f \in B_{p,q}^s, \|f\|_{B_{p,q}^s} \leq A\},$$

of radius $A > 0$ with $1 \leq p, q \leq \infty$ and $s' = s + 1/2 - 1/p$, one has

$$B_{p,q}^s(A) = \left\{ f \in L^p([0, 1]) : \left(\sum_{k=0}^{2^m-1} |a_{mk}|^p \right)^{1/p} + \left(\sum_{j=m}^{\infty} 2^{js'q} \left(\sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{q/p} \right)^{1/q} \leq A \right\}, \quad (3.1)$$

with respective sum(s) replaced by maximum if $p = \infty$ and/or $q = \infty$ (see, e.g., Johnstone *et. al* (2004)). We study below the minimax L^2 -risk over Besov balls $B_{p,q}^s(A)$ defined as

$$R_\varepsilon(B_{p,q}^s(A)) = \inf_{\tilde{f}} \sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\tilde{f} - f\|^2,$$

where the infimum is taken over all possible square-integrable estimators \tilde{f} of f based on y from model (1.1).

In what follows, we use the symbol C for a generic positive constant, which may take different values at different places and is independent of the noise level ε . The following statement provides the asymptotic minimax lower bounds for the L^2 -risk.

Theorem 1 *Let $1 \leq p, q \leq \infty$ and $s > \max(1/p, 1/2)$. Then, under Assumptions (A1)–(A3), as $\varepsilon \rightarrow 0$,*

$$R_\varepsilon(B_{p,q}^s(A)) \geq C \Delta(\varepsilon), \quad (3.2)$$

where

$$\Delta(\varepsilon) = \begin{cases} A^{\frac{2(\alpha+\beta)}{2s'+\alpha+\beta}} \varepsilon^{\frac{2s'}{2s'+\alpha+\beta}} & \text{if } 2s(\alpha-1) \geq (\beta+1)(1-2/p), \\ A^{\frac{2(\beta+1)}{2s+\beta+1}} \varepsilon^{\frac{2s}{2s+\beta+1}} & \text{if } 2s(\alpha-1) < (\beta+1)(1-2/p), \end{cases} \quad (3.3)$$

Remark 2 (Convergence rates) As we shall show below, the minimax global convergence rates in Theorem 1 are attainable up to a logarithmic factor. The rates are determined by the interaction of four parameters, s, p, α and β . Parameters s and p describe, respectively, smoothness and spatial homogeneity of the unknown function f , while β and α , defined in (2.6), are referred to as degrees of ill-posedness and spatial inhomogeneity of operator Q . If the value of α is large, in particular, $2s\alpha > 2s' + \beta(1 - 2/p)$, we say that operator Q is *strongly inhomogeneous* while in the case when $2s\alpha < 2s' + \beta(1 - 2/p)$ we call operator Q *weakly inhomogeneous*. The case when $2s\alpha = 2s' + \beta(1 - 2/p)$ is referred to as *moderately inhomogeneous*. Observe that in the weakly inhomogeneous case, spatial inhomogeneity of operator Q does not affect convergence rate which is determined entirely by the degree of ill-posedness β . On the other hand, for large values of α , convergence rate is significantly affected by the degree of spatial inhomogeneity α of Q .

4 Estimation strategies in the presence of a singularity

To be more specific, we consider a periodized version of the wavelet basis on the unit interval

$$\{\varphi_{mk}, \psi_{jk} : j \geq m, k = 0, 1, \dots, 2^j - 1\}, \quad (4.1)$$

where $\varphi_{mk}(x) = 2^{m/2}\varphi(2^m x - k)$, $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$, $x \in [0, 1]$. Note that the latter requires that the resolution level m is high enough, in particular, $m \geq m_1$, where m_1 is such that

$$2^{m_1} > \max(\mathcal{L}_{\varphi^*}, \mathcal{L}_{\psi^*}). \quad (4.2)$$

Here, \mathcal{L}_{φ^*} and \mathcal{L}_{ψ^*} are the lengths of supports of the mother and father wavelets, φ^* and ψ^* , that generate periodized wavelet basis. Then, for any $m \geq m_1$, the set (4.1) forms an orthonormal wavelet basis for $L^2([0, 1])$, and, hence, any $f \in L^2([0, 1])$, can be expanded using formula (2.5). Under Assumptions (A1), (A2) and (A4), one can construct unbiased estimators of coefficients a_{mk} and b_{jk}

$$\hat{a}_{mk} = \langle y, t_{m,k} \rangle, \quad \hat{b}_{jk} = \langle y, u_{j,k} \rangle. \quad (4.3)$$

If $k \in K_{0m}^c$ and $k \in K_{1j}^c$, respectively, then estimators \hat{a}_{mk} and \hat{b}_{jk} have finite variances

$$\text{Var}(\hat{a}_{mk}) \asymp \lambda_{m,k}^{-2}, \quad k \in K_{0m}^c, \quad \text{Var}(\hat{b}_{jk}) \asymp \lambda_{j,k}^{-2}, \quad k \in K_{1j}^c, \quad (4.4)$$

and have infinite variances otherwise. In order to account for the latter, for any $m \geq m_1$, we partition f into the sum of singularity-affected and singularity-free parts

$$f(x) = f_{0,m}(x) + f_{c,m}(x), \quad x \in [0, 1],$$

where

$$f_{0,m}(x) = \sum_{k \in K_{0m}} a_{mk} \varphi_{mk}(x) + \sum_{j=m}^{\infty} \sum_{k \in K_{1j}} b_{jk} \psi_{jk}(x), \quad x \in [0, 1], \quad (4.5)$$

$$f_{c,m}(x) = \sum_{k \in K_{0m}^c} a_{mk} \varphi_{mk}(x) + \sum_{j=m}^{\infty} \sum_{k \in K_{1j}^c} b_{jk} \psi_{jk}(x), \quad x \in [0, 1]. \quad (4.6)$$

We then construct estimators $\hat{f}_{0,m}$ and $\hat{f}_{c,m}$ of $f_{0,m}$ and $f_{c,m}$, respectively, and estimate f by a hybrid estimator

$$\hat{f}_m(x) = \hat{f}_{0,m}(x) + \hat{f}_{c,m}(x), \quad x \in [0, 1]. \quad (4.7)$$

In particular, we shall use a linear estimator with the resolution level m estimated from the data as $\hat{f}_{0,m}$ and a nonlinear block thresholding wavelet estimator as $\hat{f}_{c,m}$ with the lowest resolution level m in $\hat{f}_{c,m}$ determined by the linear part $\hat{f}_{0,m}$. In what follows, we shall consider estimation of $f_{0,m}$ and $f_{c,m}$ separately.

First, we construct a block thresholding wavelet estimator $\hat{f}_{c,m}$ of $f_{c,m}$. For this purpose, we divide the wavelet coefficients at each resolution level into $l_{j\varepsilon}^L$ blocks of length $\ln(\varepsilon^{-1})$ to the left of $(k_{0j} - D)$ and $l_{j\varepsilon}^R$ blocks to the right of $(k_{0j} + D)$, where

$$l_{j\varepsilon}^L = (k_{0j} - D) / \ln(\varepsilon^{-1}), \quad l_{j\varepsilon}^R = (2^j - D - 1 - k_{0j}) / \ln(\varepsilon^{-1}), \quad l_{j\varepsilon} = \max(l_{j\varepsilon}^L, l_{j\varepsilon}^R). \quad (4.8)$$

Define blocks U_{jl}^L and U_{jl}^R of indices k to the left of $(k_{0j} - D)$ and to the right of $(k_{0j} + D)$, respectively, as

$$U_{jl}^L = \{k : k_{0j} - D - l \ln(\varepsilon^{-1}) < k \leq k_{0j} - D - (l-1) \ln(\varepsilon^{-1})\}, \quad l \in U_j^L, \quad (4.9)$$

$$U_{jl}^R = \{k : k_{0j} + D + (l-1) \ln(\varepsilon^{-1}) < k \leq k_{0j} + D + l \ln(\varepsilon^{-1})\}, \quad l \in U_j^R, \quad (4.10)$$

where

$$U_j^L = \{l : 1 \leq l \leq l_{j\varepsilon}^L\}, \quad U_j^R = \{l : 1 \leq l \leq l_{j\varepsilon}^R\}, \quad U_j = U_j^L \cup U_j^R. \quad (4.11)$$

To simplify the narrative, we shall write $l \in U_j$ and $k \in U_{jl}$ without a specific reference whether a block lies to the right or to the left of k_{0j} . Denote

$$B_{jl} = \sum_{k \in U_{jl}} b_{jk}^2, \quad \hat{B}_{jl} = \sum_{k \in U_{jl}} \hat{b}_{jk}^2, \quad (4.12)$$

$$R_{jl\varepsilon} = \frac{\varepsilon \ln(\varepsilon^{-1}) 2^{j(\alpha+\beta)}}{|D + (l-1) \ln(\varepsilon^{-1})|^\alpha} \asymp \varepsilon \sum_{k \in U_{jl}} \lambda_{j,k}^{-2}. \quad (4.13)$$

For any $m \geq m_1$, estimate $f_{c,m}$ by

$$\hat{f}_{c,m}(x) = \sum_{k \in K_{0m}^c} \hat{a}_{mk} \varphi_{mk}(x) + \sum_{j=m}^{J-1} \sum_{l \in U_j} \sum_{k \in U_{jl}} \hat{b}_{jk} \mathbb{I}(\hat{B}_{jl} \geq \tau^2 R_{jl\varepsilon}) \psi_{jk}(x), \quad (4.14)$$

where $\mathbb{I}(\Omega)$ is the indicator function of the set Ω , the value of τ will be defined later and

$$2^J = \varepsilon^{-\frac{2}{\alpha+\beta+2}}. \quad (4.15)$$

Now, consider estimation of the singularity-affected part. Since the estimators \hat{a}_{mk} of a_{mk} , given in (4.3), have infinite variances when $k \in K_{0m}$, we estimate those coefficients by solving a system of linear equations. Denote $w_{m,k} = Q\varphi_{mk}$ and observe that, for a given m , $m_1 \leq m \leq J-1$, one has $f = f_m + R_m$. Here

$$f_m = \sum_{k \in K_{0m}} a_{mk} \varphi_{mk} + \sum_{k \in K_{0m}^c} a_{mk} \varphi_{mk}, \quad R_m = \sum_{j=m}^{\infty} \sum_{k=0}^{2^j-1} b_{jk} \psi_{jk}, \quad (4.16)$$

and, hence,

$$Qf = \sum_{k \in K_{0m}} a_{mk} w_{m,k} + \sum_{k \in K_{0m}^c} a_{mk} w_{m,k} + QR_m. \quad (4.17)$$

Taking scalar products of both sides of (4.17) with $w_{m,l}$, $l \in K_{0m}$, obtain

$$\langle w_{m,l}, Qf \rangle = \sum_{k \in K_{0m}} a_{mk} \langle w_{m,l}, w_{m,k} \rangle + \sum_{k \in K_{0m}^c} a_{mk} \langle w_{m,l}, w_{m,k} \rangle + \langle w_{m,l}, QR_m \rangle, \quad l \in K_{0m}. \quad (4.18)$$

Introduce matrices $\mathbf{A}^{(m)}$ and $\mathbf{B}^{(m)}$ and vectors $\mathbf{c}^{(m)}$, $\hat{\mathbf{c}}^{(m)}$, $\mathbf{r}^{(m)}$, $\mathbf{z}^{(m)}$, $\mathbf{h}^{(m)}$ and $\hat{\mathbf{h}}^{(m)}$ with elements

$$A_{lk}^{(m)} = \langle w_{m,l}, w_{m,k} \rangle, \quad k, l \in K_{0m}, \quad B_{lk}^{(m)} = \langle w_{m,l}, w_{m,k} \rangle, \quad k \in K_{0m}^c, l \in K_{0m}, \quad (4.19)$$

$$c_l^{(m)} = \langle w_{m,l}, Qf \rangle, \quad l \in K_{0m}, \quad \hat{c}_l^{(m)} = \langle w_{m,l}, y \rangle, \quad l \in K_{0m}, \quad (4.20)$$

$$r_l^{(m)} = \langle w_{m,l}, QR_m \rangle, \quad l \in K_{0m}, \quad z_k^{(m)} = a_{mk}, \quad k \in K_{0m}, \quad (4.21)$$

$$h_k^{(m)} = a_{mk}, \quad k \in K_{0m}^c, \quad \hat{h}_k^{(m)} = \hat{a}_{mk} = \langle y, t_{m,k} \rangle, \quad k \in K_{0m}^c, \quad (4.22)$$

where \hat{a}_{mk} , $k \in K_{0m}$, are defined in (4.3). Then, one can re-write an exact system of linear equations (4.18) as $\mathbf{c}^{(m)} = \mathbf{A}^{(m)} \mathbf{z}^{(m)} + \mathbf{B}^{(m)} \mathbf{h}^{(m)} + \mathbf{r}^{(m)}$ and obtain its approximate version

$$\hat{\mathbf{c}}^{(m)} = \mathbf{A}^{(m)} \hat{\mathbf{z}}^{(m)} + \mathbf{B}^{(m)} \hat{\mathbf{h}}^{(m)}. \quad (4.23)$$

Since matrix $\mathbf{A}^{(m)}$ is a nonnegative definite matrix of a finite size, in order to guarantee that it is nonsingular, it is sufficient to impose the following almost trivial assumption:

(A5) Functions $w_{m,k} = Q\varphi_{mk}$, $k \in K_{0m}$, are linearly independent.

Under Assumption (A5), one has

$$\mathbf{z}^{(m)} = (\mathbf{A}^{(m)})^{-1} \left(\mathbf{c}^{(m)} - \mathbf{B}^{(m)} \mathbf{h}^{(m)} - \mathbf{r}^{(m)} \right), \quad \hat{\mathbf{z}}^{(m)} = (\mathbf{A}^{(m)})^{-1} \left(\hat{\mathbf{c}}^{(m)} - \mathbf{B}^{(m)} \hat{\mathbf{h}}^{(m)} \right). \quad (4.24)$$

Finally, for a given m , we set $\hat{a}_{mk} = \hat{z}_k^{(m)}$, $k \in K_{0m}$, and estimate $f_{0,m}$ by the following wavelet linear estimator

$$\hat{f}_{0,m}(x) = \sum_{k \in K_{0m}} \hat{a}_{mk} \varphi_{mk}(x), \quad x \in [0, 1]. \quad (4.25)$$

Remark 3 (Relation to nonparametric regression estimation based on spatially inhomogeneous data) We need to touch upon relationship between the present paper and the paper by Antoniadis, Pensky and Sapatinas (2012) which considered nonparametric regression estimation based on irregularly spaced data, in particular, in the case when design density has zeros. The latter problem is the well known formulation and has been studied extensively by many authors, including the case of the design density with zeros (see e.g., Gaïffas (2005, 2006, 2007, 2009)). On the other hand, the present paper introduces a completely novel notion of a spatially inhomogeneous linear inverse problem, discusses the best precision with which unknown function f can be recovered in the case when f itself is possibly spatially inhomogeneous and suggests estimation algorithm which allows to attain this precision in an adaptive fashion.

The common ground between the two papers is that regression estimation with vanishing design density is indeed an example of a spatially inhomogeneous ill-posed problem and can be considered as a trivial case of deconvolution with spatially inhomogeneous design in Section 7.2 with Q being an identity operator. For this reason, the hybrid estimator was proposed in Antoniadis, Pensky and Sapatinas (2012), although, due to the fact that in the regression set up one observes function f directly, construction of the hybrid estimator is much more involved in the case of an inverse problem than in the case of nonparametric regression. In addition, the present paper provides the implementation of the hybrid estimator and studies its performance via simulations which has never been done previously since Antoniadis, Pensky and Sapatinas (2012) considered only theoretical construction of the hybrid estimator.

5 Risks of the estimators of the singularity-free and the singularity-affected parts.

In this section we shall provide asymptotic expressions for the risks of estimators (4.14) and (4.25) when the lowest resolution level m in both of them is a fixed, non-random quantity possibly dependent on ε : $m = m(\varepsilon)$.

Let us first construct an asymptotic upper bound for the singularity-free portion (4.14) of the estimator. Denote

$$\lambda_m^{-2} = \sum_{k \in K_{0m}^c} \lambda_{m,k}^{-2} \quad (5.1)$$

and observe that, under condition (2.6), there exist positive constants $C_{\lambda\alpha 0}$ and $C_{\lambda\alpha}$ independent of m such that

$$C_{\lambda\alpha 0} 2^{m(\beta + \max(1, \alpha))} m^{\mathbb{I}(\alpha=1)} \leq \lambda_m^{-2} \leq C_{\lambda\alpha} 2^{m(\beta + \max(1, \alpha))} m^{\mathbb{I}(\alpha=1)}. \quad (5.2)$$

Lemma 1 Let $1 \leq p, q \leq \infty$, $s \geq \max(1/2, 1/p)$, $s^* = \min(s, s')$ and Assumptions (A1)-(A4) hold. Let $\hat{f}_{c,m}$ be given by (4.14) where the non-random quantity $m = m(\varepsilon)$ is such that $m_1 \leq m \leq J-1$, with J defined in (4.15). If

$$\tau^2 = 4C_u C_\lambda (\sqrt{2}\chi + 1)^2 \quad \text{and} \quad \chi^2 \geq \frac{4(\beta + \max(1, \alpha))}{2 + \alpha + \beta}, \quad (5.3)$$

where C_u and C_λ are defined in (2.2) and (2.6), respectively, then, as $\varepsilon \rightarrow 0$,

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{c,m} - f_{c,m}\|^2 \leq C (\lambda_m^{-2} \varepsilon + \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho). \quad (5.4)$$

Here $\Delta(\varepsilon)$ is defined in (3.3),

$$\rho = \begin{cases} \mathbb{I}(2s(\alpha - 1) = (\beta + 1)(1 - 2/p)) + \frac{(1-\alpha)(2-p)}{2-\alpha p} \mathbb{I}(\alpha < 1) \mathbb{I}(p < 2), & \text{if } \alpha \neq 1, \\ (2s^* + \beta + 1)^{-1} 2s^*, & \text{if } \alpha = 1. \end{cases} \quad (5.5)$$

and \mathbb{I} is the indicator function. Moreover, as $\varepsilon \rightarrow 0$,

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{c,m} - f_{c,m}\|^4 = o(\varepsilon^{-2}). \quad (5.6)$$

Now, we find upper bounds for the singularity-affected portion of the estimator $\hat{f}_{0,m}(x)$. Recall that $w_{m,k} = Q\varphi_{mk}$ and let ρ_m be such that

$$C_{w1}\rho_m \leq \|w_{m,k}\| \leq C_{w2}\rho_m \quad \text{if } k \in K_{0m}. \quad (5.7)$$

Note that, since the set K_{0m} contains at most $(2D_0)$ indices, ρ_m satisfying condition (5.7) can always be found. The advantage of using the system of equations (4.23) rests upon the fact that matrix $\mathbf{A}^{(m)}$ is a finite dimensional positive definite matrix with all eigenvalues of order ρ_m^2 . In particular,

$$\|\mathbf{A}^{(m)}\| \leq C_{A1}\rho_m^2, \quad \|(\mathbf{A}^{(m)})^{-1}\| \leq C_{A2}\rho_m^{-2} \quad (5.8)$$

for some positive constants C_{A1} and C_{A2} independent of m , as it is shown in the proof of the following lemma which states the rate of convergence of the singularity affected portion of the estimator.

Lemma 2 Let $1 \leq p, q \leq \infty$ and $s \geq \max(1/2, 1/p)$. Let Assumptions (A1)-(A5) hold and there exists $C_{\rho\lambda}$, $0 < C_{\rho\lambda} < \infty$, independent of m , such that

$$\rho_m^2 \geq C_{\rho\lambda} \lambda_m^2, \quad (5.9)$$

where λ_m and ρ_m are defined in (5.1) and (5.7), respectively. Let $\alpha \geq 1$ and also

$$\rho_m^{-4} \max_{l \in K_{0m}} \sum_{k \in K_{0m}^c} \lambda_{m,k}^{-2} \langle w_{m,l}, w_{m,k} \rangle^2 \leq K_1 \lambda_m^{-2} \quad (5.10)$$

$$\max_{l \in K_{0m}} \sum_{j=m}^{\infty} \sum_{k=0}^{2^j-1} |\langle w_{m,l}, v_{j,k} \rangle| \leq K_2 \rho_m^2, \quad (5.11)$$

for some absolute constants K_1 and K_2 independent of m . If estimator $\hat{f}_{0,m}$ of $f_{0,m}$ is given by (4.25), then for any m , $m_1 \leq m \leq J-1$, and some constant C independent of m and ε , as $\varepsilon \rightarrow 0$, one has

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{0,m} - f_{0,m}\|^2 \leq C \left(\varepsilon \lambda_m^{-2} + 2^{-2ms'} \right), \quad (5.12)$$

and, moreover,

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{0,m} - f_{0,m}\|^4 = o(\varepsilon^{-2}). \quad (5.13)$$

Note that in order $\hat{f}_m = \hat{f}_{c,m} + \hat{f}_{0,m}$ estimates f , one needs to start the estimator $\hat{f}_{c,m}$ in (4.14) at exactly the same resolution level at which the linear estimator $\hat{f}_{0,m}$ in (4.25) is constructed. Thus, the choice of the lowest resolution level in (4.14) is driven by the choice of m in (4.25).

Let m_0 be such that

$$2^{m_0} = \left(\varepsilon [\ln(\varepsilon^{-1})]^{\mathbb{I}(\alpha=1)} \right)^{-\frac{1}{2s'+\alpha+\beta}}, \quad (5.14)$$

so that, for $\alpha \geq 1$, one has

$$\varepsilon \lambda_{m_0}^{-2} + 2^{-2m_0 s'} \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho. \quad (5.15)$$

The following statement delivers the total squared risk of the estimator (4.7) of f if $DD_0 > 0$.

Theorem 2 *Let $1 \leq p, q \leq \infty$ and $s \geq \max(1/2, 1/p)$. Let conditions (5.9)–(5.11) and Assumptions (A1)–(A5) hold with $\alpha \geq 1$ and $DD_0 > 0$ in Assumptions (A2) and (A4). Consider estimator (4.7) of f where $\hat{f}_{c,m}$ and $\hat{f}_{0,m}$ are given by formulae (4.14) and (4.25), respectively. Let $m = m_0$ where m_0 is defined in (5.14), J be defined in (4.15) and let positive constants τ and χ satisfy condition (5.3). Then, for some constant C independent of ε , as $\varepsilon \rightarrow 0$, one has*

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{m_0} - f\|^2 \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho \quad (5.16)$$

where $\Delta(\varepsilon)$ and ρ are defined in (3.3) and (5.5), respectively.

Validity of Theorem 2 follows directly from Lemmas 1 and 2 and inequality (5.15).

Remark 4 (Optimality) Note that in (5.5), $\rho = 0$ unless $2s(\alpha - 1) = (\beta + 1)(1 - 2/p)$ or $\alpha = 1$ or $\alpha < 1$ and $p \leq 2$. The latter shows that the lower bounds for the risk in Theorem 1 cannot be made tighter, at least, in the case when $\alpha > 1$. Theorems 1 and 2 and Corollary 1 demonstrate that the estimator (4.7) attains the asymptotically optimal (in the minimax sense) convergence rates if $2s(\alpha - 1) \neq (\beta + 1)(1 - 2/p)$ and $\alpha > 1$ or if $\alpha < 1$ and $p > 2$. Otherwise, estimator (4.7) is asymptotically near-optimal up to a logarithmic factor.

Note that the value of m_0 depends on the unknown parameters s and p of the Besov space, hence, in general, estimator \hat{f}_{m_0} in (5.16) is not adaptive. However, if $D = D_0 = 0$, then $\hat{f}_{0,m} = \hat{f}_{0,m} \equiv 0$, $\hat{f} = \hat{f}_{c,m}$ and one can choose $m = m_1$ in $\hat{f}_{c,m}$ using formula (4.2), so that the estimator is adaptive. In this case, convergence rates of \hat{f} are given entirely by Lemma 1. In particular, the following Corollary is valid.

Corollary 1 *Let $D = D_0 = 0$ and assumptions of Lemma 1 hold. Consider estimator $\hat{f}_m = \hat{f}_{c,m}$ given by (4.14) with $m = m_1$. Then then, as $\varepsilon \rightarrow 0$,*

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{m_1} - f\|^2 \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho \quad (5.17)$$

where $\Delta(\varepsilon)$ and ρ are defined in (3.3) and (5.5), respectively.

If $DD_0 > 0$, then it is necessary to construct an adaptive estimator of f . Note that this is not an easy task. Expanding the system of equations in (4.18) so that it includes not only the scaling but also the wavelet coefficients will compromise uniformity of eigenvalues of matrix $\mathbf{A}^{(m)}$ (see (5.8)) which are ensured by positive-definiteness and finite size of $\mathbf{A}^{(m)}$. On the other hand, introducing a penalty on the solution does not help either since, for any m , the system involves the unknown bias term $\mathbf{r}^{(m)}$. For this reason, in order to choose parameter m , we apply Lepskii method since it allows to eliminate the bias inherent to the system of equations (4.23).

6 Adaptive estimation in the presence of singularity

In order to construct an adaptive estimator of f in the presence of singularity, we shall use the technique of optimal tuning parameter selection pioneered by Lepski (1990, 1991) and further exploited in Lepski and Spokoiny (1997) and Lepski *et al.* (1997). The idea behind this technique is to construct estimators for various values of the tuning parameter in question (m , in our case), and then choose an optimal value of the tuning parameter by regulating the differences between the estimators constructed with different values of the parameter.

In particular, for various values of m , we construct versions of the system of equations, obtain values of $\hat{\mathbf{z}}^{(m)}$ in (4.24) and use them as $\hat{a}_{mk} = \hat{z}_k^{(m)}$, $k \in K_{0m}$, in (4.25). Then, for various values of m , we obtain estimators \hat{f}_m of f using formula (4.7) where $\hat{f}_{0,m}$ and $\hat{f}_{c,m}$ are of the forms (4.25) and (4.14), respectively, and m is the lowest resolution level of $\hat{f}_{c,m}$. After that, we choose the “best possible” resolution level \hat{m} and consider estimator $\hat{f}_{\hat{m}}$ as the final estimator. The choice of the resolution level \hat{m} is driven by the singularity-affected portion of f rather than the zero-free portion as it is described below.

For any resolution level $m > 0$, we define a neighborhood Ω_m of x_0 as

$$\Omega_m = \{x : \min(L_\varphi - D_0, L_\psi - D) < 2^m(x - x_0) < \max(U_\varphi + D_0, U_\psi + D)\} \quad (6.1)$$

where $\text{supp } \varphi = (L_\varphi, U_\varphi)$ and $\text{supp } \psi = (L_\psi, U_\psi)$. Observe that Ω_m is designed so that $\text{supp}(f_{0,m}) \subseteq \Omega_m$, $\text{supp}(\hat{f}_{0,m}) \subseteq \Omega_m$ and $\Omega_j \subset \Omega_m$ if $j > m$.

Choose $m = \hat{m}$ such that $m_1 \leq m \leq J - 1$, where J is defined in (4.15) and

$$\hat{m} = \min \left\{ m : \|(\hat{f}_m - \hat{f}_j)\mathbb{I}(\Omega_m)\|^2 \leq \kappa^2 \varepsilon \ln(\varepsilon^{-1}) \lambda_j^{-2} \text{ for all } j, m \leq j \leq J - 1 \right\}, \quad (6.2)$$

where $\kappa > 0$ is a constant to be defined below.

The construction of \hat{m} is based on the following idea. Note that when $m = \hat{m} \leq m_0$, then one has

$$\mathbb{E}\|\hat{f}_{\hat{m}} - f\|^2 \leq 2 \left[\mathbb{E}\|\hat{f}_{\hat{m}} - \hat{f}_{m_0}\|^2 + \mathbb{E}\|\hat{f}_{m_0} - f\|^2 \right]. \quad (6.3)$$

The first component in (6.3) is small due to definition of the resolution level \hat{m} while the second component is calculated at the optimal resolution level m_0 and, hence, tends to zero at the optimal convergence rate (up to a logarithmic factor). On the other hand, if $m = \hat{m} > m_0$, then there exists $j > m$ such that $\|(\hat{f}_m - \hat{f}_j)\mathbb{I}(\Omega_m)\|^2 > \kappa^2 \varepsilon \ln(\varepsilon^{-1}) \lambda_j^{-2}$. The following Lemma shows that, provided κ is large enough, the probability of this event is infinitesimally small.

Lemma 3 *Let m_0 and \hat{m} be given by expressions (5.14) and (6.2), respectively, and conditions of Theorem 2 hold. Let $C_\kappa = 2^{11} D_0^2 C_{A2}^2 \max(C_{\rho\lambda}^{-1} C_{w2}^2, C_t K_1)$ where constants C_{A2} , $C_{\rho\lambda}$, C_{w2} , C_t and K_1 are defined in (5.8), (5.9), (5.7), (2.4) and (5.10), respectively. If*

$$\kappa \geq \max(d^2 C_\kappa, 2^6(d^2 + 1)C_t), \quad \chi^2 \geq d^2 + 2/(2 + \alpha + \beta), \quad (6.4)$$

then, as $\varepsilon \rightarrow 0$,

$$\mathbb{P}(\hat{m} > m_0) \leq C\varepsilon^{d^2}. \quad (6.5)$$

Lemma 3 confirms that indeed $m = \hat{m}$ can be chosen as the lowest resolution level in the nonlinear portion of the estimator, so that we estimate f by

$$\hat{f}(x) = \hat{f}_{0,\hat{m}}(x) + \hat{f}_{c,\hat{m}}(x), \quad x \in [0, 1], \quad (6.6)$$

where $\hat{f}_{0,m}(x)$ and $\hat{f}_{c,m}(x)$ are defined in (4.25) and (4.14), respectively. The following statement confirms that the wavelet nonlinear estimator \hat{f} given by (6.6) indeed attains (up to a logarithmic factor) the asymptotic minimax lower bounds obtained in Theorem 1.

Theorem 3 Let $1 \leq p, q \leq \infty$ and $s \geq \max(1/2, 1/p)$. Let conditions (5.9)–(5.11) and Assumptions (A1)–(A5) hold with $\alpha \geq 1$ and $DD_0 > 0$ in Assumptions (A2) and (A4). Consider estimator (4.7) of f where $\hat{f}_{c,m}$ and $\hat{f}_{0,m}$ are given by formulae (4.14) and (4.25), respectively. Let $m = \hat{m}$ where \hat{m} is defined in (6.2). Let J be defined in (4.15) and κ and χ be such that

$$\kappa \geq \max(4C_\kappa, 320C_t), \quad \chi^2 \geq 4 + 2/(2 + \alpha + \beta), \quad (6.7)$$

where C_κ and C_t are defined in Lemma 3 and formula (2.4), respectively. Then, for some constant C independent of ε , as $\varepsilon \rightarrow 0$, one has

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{\hat{m}} - f\|^2 \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^{1+\mathbb{I}(\alpha=1)} \quad (6.8)$$

where $\Delta(\varepsilon)$ is defined in (3.3).

Remark 5 (Logarithmic factor in convergence rates) Observe that in Theorem 2 convergence rates are sharp if $\alpha > 1$ or $p \geq 2$ unless $2s(\alpha - 1) = (\beta + 1)(1 - 2/p)$. However, in Theorem 3 the risk of the adaptive estimator is always within a logarithmic factor $[\ln(\varepsilon^{-1})]^{1+\mathbb{I}(\alpha=1)}$ of the minimax risk. The latter is due to application of Lepskii method. Note that in spite of the fact that we are using mean squared error, Lepskii method is applied locally and hence leads to an extra log-factor in the risk, as it usually happens with application of Lepskii method to pointwise estimation.

7 Examples

7.1 Deconvolution with a spatially inhomogeneous kernel

Consider problem (1.1) with operator Q of the form

$$(Qf)(x) = \mu(x) \int_0^1 q(x-t)f(t)dt, \quad x \in [0, 1], \quad (7.1)$$

where functions $\mu(x)$, $q(x)$ and $f(x)$ in the right-hand side of equation (7.1) are periodic and both $q(x)$ and $\mu(x)$ are completely known. In this case, problem (1.1) is known to be equivalent to the following statistical problem

$$y_i = \mu(i/n) \int_0^1 q(i/n-t)f(t)dt + \sigma \xi_i, \quad i = 1, \dots, n, \quad (7.2)$$

where ξ_i is a white Gaussian noise and $\varepsilon = \sigma^2/n$. Equation of the form (7.1) can appear when one observes a convolution $Y(x)$ of the known kernel q with the unknown function of interest f and a known heteroscedastic noise $\sqrt{\varepsilon}\gamma(x)W(x)$, so that $\mu(x) = [\gamma(x)]^{-1}$ and $Y(x) = \gamma(x)y(x) = y(x)/\mu(x)$. In this case, equation (7.2) takes the form

$$Y_i = \int_0^1 q(i/n-t)f(t)dt + \sigma \gamma(i/n)\xi_i, \quad i = 1, \dots, n. \quad (7.3)$$

If $\mu(x)$ is uniformly bounded above and below, in principle, spatial inhomogeneity of operator Q in (7.1) can be ignored. Below we consider the case when the former is not true since $\mu(x)$ vanishes at some point $x_0 \in (0, 1)$, in particular,

$$C_{\mu 1}|x - x_0|^\alpha \leq \mu^2(x) \leq C_{\mu 2}|x - x_0|^\alpha \quad (7.4)$$

for some for some positive constants α , $C_{\mu 1}$ and $C_{\mu 2}$ independent of x_0 and x . Therefore, the version of the problem studied in the present paper can be described as locally extreme noise which arises when the degree of spatial inhomogeneity is high.

Direct calculations show that

$$(Q^*h)(z) = \int_0^1 q(x-z)h(x)\mu(x)dx.$$

Hence, $u_{j,k}$ and $t_{m,k}$ are solutions of the equations $Q^*u_{j,k} = \psi_{jk}$ and $Q^*t_{m,k} = \varphi_{mk}$. Consider $\omega \in \mathbb{Z}$ and let $e_\omega(t) = e^{i2\pi\omega t}$, $t \in [0,1]$, be the periodic Fourier basis on $[0,1]$. Denote $U_{jk}(x) = u_{j,k}(x)\mu(x)$ and, similarly, $T_{mk}(x) = t_{m,k}(x)\mu(x)$ and introduce Fourier coefficients $q_\omega = \langle q, e_\omega \rangle$, $U_{jk\omega} = \langle U_{jk}, e_\omega \rangle$, $T_{mk\omega} = \langle T_{mk}, e_\omega \rangle$, $\psi_{jk\omega} = \langle \psi_{jk}, e_\omega \rangle$ and $\varphi_{mk\omega} = \langle \varphi_{mk}, e_\omega \rangle$. Then, $U_{jk\omega} = [\bar{q}_\omega]^{-1} \psi_{jk\omega}$, $T_{mk\omega} = [\bar{q}_\omega]^{-1} \varphi_{mk\omega}$ and

$$U_{jk}(x) = \sum_{\omega \in \mathbb{Z}} [\bar{q}_\omega]^{-1} \psi_{jk\omega} e_\omega(x), \quad T_{mk}(x) = \sum_{\omega \in \mathbb{Z}} [\bar{q}_\omega]^{-1} \varphi_{mk\omega} e_\omega(x), \quad (7.5)$$

where \bar{q}_ω is the complex conjugate of q_ω . Moreover, estimators \hat{a}_{mk} and \hat{b}_{jk} in (4.3) can be constructed using Fourier wavelet transform suggested in Johnstone *et al.* (2004). Indeed, if $Y(x) = y(x)/\mu(x)$ and Y_ω are Fourier coefficients of function $Y(x)$, then

$$\hat{b}_{jk} = \langle y, u_{j,k} \rangle = \langle Y, U_{jk} \rangle = \sum_{\omega \in \mathbb{Z}} [\bar{q}_\omega]^{-1} \psi_{jk\omega} \bar{Y}_\omega, \quad \hat{a}_{mk} = \sum_{\omega \in \mathbb{Z}} [\bar{q}_\omega]^{-1} \varphi_{mk\omega} \bar{Y}_\omega. \quad (7.6)$$

In the case of the statistical experiment described in formula (7.2), Fourier coefficients are replaced by the discrete Fourier transform.

Note that application of the wavelet-vaguellete methodology to deconvolution with heteroscedastic noise (7.3) is very reasonable. Really, if noise level $\gamma(x)$ is such that $\int \gamma^2(x)dx < \infty$, then formula (7.6) implies that wavelet coefficients are estimated using Fourier transform of the measured signal Y in (7.3) and then thresholded taking into account the local noise level. Indeed, it is easy to observe that $\lambda_{j,k}$ in (7.10) is such that $\lambda_{j,k}^{-2} \asymp \|U_{jk}\|^2 \int \gamma^2(x) \mathbb{I}(x \in \text{supp } U_{jk}) dx$. If $\int \gamma^2(x)dx = \infty$ in the vicinity of some point x_0 , the natural strategy suggested above cease working and one needs another means for estimating scaling and wavelet coefficients in the neighborhood of x_0 . This is the situation when one has to apply the hybrid estimator constructed in Sections 4 and 6. In Section 8 we describe in detail how this task can be accomplished. If function $\mu(x)$ is not completely known and is estimated from data, matrices $\mathbf{A}^{(m)}$ and $\mathbf{B}^{(m)}$ as well as vector $\hat{\mathbf{c}}^{(m)}$ will be subjected to additional errors which have to be accounted for by using regularization techniques designed for the inverse problems with errors in the operator (see, e.g., Engl, Hanke and Neubauer (2000) and Hoffmann and Reiss (2008)).

In order to find $\lambda_{j,k} \asymp \|v_{j,k}\|$ in Assumption (A1) and verify Assumptions (A1)–(A5), we impose the following conditions on the kernel q and mother and father wavelets ψ and φ .

(E1) Kernel $q(x)$ is $(r-2)$ times continuously differentiable on $[0,1]$ and $r_1 > r \geq 1$ times differentiable outside the neighborhood of jump discontinuities of $q^{(r-1)}$ with $q^{(r)}$ and $q^{(r_1)}$ uniformly bounded. The value $r = 1$ corresponds to the case when q itself has jump discontinuities.

(E2) Fourier coefficients q_ω of q are such that $C_{q1}(|\omega| + 1)^{-r} \leq |q_\omega| \leq C_{q2}(|\omega| + 1)^{-r}$ for some positive constants C_{q1} and C_{q2} independent of ω .

(E3) Let ψ be r_0 -regular, r_0 times continuously differentiable wavelet function with the bounded support, where $r_0 > \max(r, r_1)$.

(E4) Kernel q is such that functions U_{jk} and T_{jk} defined in (7.5) have bounded supports of the lengths proportional to 2^{-j} and centered at $2^{-j}k$: $\text{supp}(U_{jk}) = (2^{-j}(k - d_U), 2^{-j}(k + d_U))$

and $\text{supp}(T_{jk}) = (2^{-j}(k - d_T), 2^{-j}(k + d_T))$.

There are many functions q satisfying conditions above, among them, for example,

$$q_1(x) = \sum_{k \in \mathbb{Z}} \exp(-\lambda|x + k|) \quad \text{and} \quad q_2(x) = \sum_{k \geq 0} \exp(-\lambda(x + k))(x + k)^N, \quad (7.7)$$

with $r = 2$ for $q_1(x)$ and $r = N + 1$ for $q_2(x)$.

Note that under Assumptions (E1)-(E4), one has $\|u_{j,k}\|^2 = \int \mu^{-2}(x) U_{jk}^2(x) dx = \infty$ if $|k - k_{0j}| < d_U$ and $\alpha \geq 1$ and

$$\begin{aligned} \|u_{j,k}\|^2 &\asymp \mu^{-2}(2^{-j}k) \|U_{jk}\|^2 \asymp |2^{-j}k - x_0|^{-\alpha} \sum_{\omega \in \mathbb{Z}} |\bar{q}_\omega|^{-2} |\psi_{jk\omega}|^2 \\ &\leq C 2^{j\alpha} [|k - k_{0j}|^\alpha + 1]^{-1} \sum_{\omega \in \mathbb{Z}} (|\omega|^{2r} + 1) |\psi_{jk\omega}|^2, \end{aligned} \quad (7.8)$$

otherwise. Due to conditions (E3) and (E4) and periodicity of ψ_{jk} , integrating by parts r times, we obtain the following expression for Fourier coefficients of ψ_{jk}

$$\psi_{jk\omega} = \int_0^1 2^{j/2} \psi(2^j x - k) e^{i2\pi\omega x} dx = 2^{j(r+1/2)} (-2\pi i \omega)^r \int_0^1 \psi^{(r)}(2^j x - k) e^{i2\pi\omega x} dx,$$

so that

$$2^{-2jr} \sum_{\omega \in \mathbb{Z}} (|\omega|^{2r} + 1) |\psi_{jk\omega}|^2 \leq C(2^{-2jr} \|\psi\|^2 + \|\psi^{(r)}\|^2).$$

Therefore,

$$\|u_{j,k}\|^2 \leq C 2^{j(\alpha+2r)} [|k - k_{0j}|^\alpha + 1]^{-1} \asymp \lambda_{j,k}^{-2}, \quad \text{if } |k - k_{0j}| \geq d_U \text{ or } \alpha < 1, \quad (7.9)$$

and $\|u_{j,k}\| = \infty$ otherwise. Similar inequality can be proved for $\|t_{m,k}\|^2$.

Now, we need to show that indeed $\lambda_{j,k}$ is defined by expression (7.9) and that, under conditions (E1)-(E4), Assumptions (A1)-(A5) hold. This is accomplished by the following proposition.

Proposition 1 *Let $\mu(x)$ satisfy condition (7.4). Then, under Assumptions (E1)–(E4) with r_1 such that $2r_1 + 1 > 2r + \alpha$, one has*

$$\lambda_{j,k}^2 \asymp 2^{-j(2r+\alpha)} [|k - k_{0j}|^\alpha + 1]. \quad (7.10)$$

Furthermore, Assumptions (A1)–(A5) hold with $D = D_0 = 0$ if $\alpha < 1$, and with $D = d_U$ and $D_0 = d_T$ if $\alpha \geq 1$. Conditions (5.9)–(5.11) of Lemma 2 are also valid.

Due to Proposition 1, all statements and constructions in Sections 3–6 can be applied to equation (1.1) with operator Q given in (7.1). In particular, Theorems 1–3 can be utilized with $\beta = 2r$.

By direct comparison with, e.g., Johnstone *et al.* (2004), one can see that if $\alpha = 0$, so that the problem is spatially homogeneous, then the rates of convergence in Theorems 1–3 coincide with the usual convergence rates exhibited in deconvolution problems with white noise.

7.2 Deconvolution with spatially inhomogeneous design

Consider the problem of deconvolution when measurements are irregularly spaced. In particular, let g be a sampling pdf with the corresponding cdf G . Due to irregular design, operator Q can be presented as

$$(Qf)(x) = \int_0^1 q(G^{-1}(x) - t)f(t)dt, \quad (7.11)$$

where where G^{-1} is the inverse of G . In this case, equation (1.1) can be viewed as an idealized version of the equation

$$y_i = \int_0^1 q(x_i - t)f(t)dt + \sigma\xi_i, \quad i = 1, \dots, n, \quad (7.12)$$

where $\varepsilon = \sigma^2/n$, ξ_i is a white Gaussian noise and observation points x_i , $i = 1, \dots, n$, are such that $G(i/n) = x_i$ and x_i 's and ξ_j 's are independent. Then, $G^{-1}(i/n) = x_i$, $i = 1, \dots, n$, and the right-hand side of (1.1) with operator Q given by (7.11) provides a continuous equivalent of the statistical problem (7.12).

In what follows, we assume that functions $q(x)$ and $f(x)$ are periodic and both $q(x)$ and $g(x)$ are completely known. In this case, the conjugate operator Q^* is of the form

$$(Q^*h)(z) = \int_0^1 q(G^{-1}(x) - z)h(x)dx.$$

It is pretty straightforward to show that $u_{j,k}(x) = U_{jk}(G^{-1}(x))$ and $t_{m,k}(x) = T_{mk}(G^{-1}(x))$ where, as before, $U_{jk}(\cdot)$ and $T_{mk}(\cdot)$ are given by formula (7.5). Wavelet coefficients \hat{b}_{jk} and \hat{a}_{mk} can also be estimated in a manner similar to Example 1. Indeed, if $Y(x) = y(G(x))$ and Y_ω are Fourier coefficients of Y , then \hat{b}_{jk} and \hat{a}_{mk} can be evaluated using formula (7.6).

If design density g is unknown, then both $g(x)$ and $G(x)$ have to be estimated from observations x_i , $i = 1, \dots, n$. The latter will lead to additional errors in estimating wavelet coefficients \hat{b}_{jk} and \hat{a}_{mk} as well as entries of matrices $\mathbf{A}^{(m)}$ and $\mathbf{B}^{(m)}$ and vector $\hat{\mathbf{c}}^{(m)}$. The issue of additional errors has to be addressed by using, for example, regularization techniques (see e.g. Engl, Hanke and Neubauer (2000) and Hoffmann and Reiss (2008)).

We assume that design density $g(x)$ has a single zero of order α at x_0 , i.e., $g(x_0 + x)|x|^{-\alpha} \rightarrow C_g$ as $x \rightarrow 0$. The latter implies that there exist some absolute constants C_{g1} and C_{g2} such that, for any x , one has

$$C_{g1}|x - x_0|^\alpha \leq g(x) \leq C_{g2}|x - x_0|^\alpha. \quad (7.13)$$

Thus, we are considering the case of extremely inhomogeneous design which can be described also as a local data loss.

In this case, under conditions (E1)-(E4), similarly to Example 1, one has $\|u_{j,k}\|^2 = \int g^{-1}(x)U_{jk}^2(x)dx$. Hence, identically to (7.9), one has

$$\begin{aligned} \|u_{j,k}\|^2 &\leq C2^{j(\alpha+2r)}|k - k_{0j}|^{-\alpha}, \quad \text{if } |k - k_{0j}| \geq d_U \text{ or } \alpha < 1, \\ \|u_{j,k}\|^2 &= \infty, \quad \text{if } |k - k_{0j}| < d_U \text{ and } \alpha \geq 1. \end{aligned}$$

Moreover, by simple modifications of the proof of Proposition 1, it easy to show that the following statement is valid.

Proposition 2 *Let $g(x)$ satisfy condition (7.13). Then, under Assumptions (E1)–(E4) with r_1 such that $2r_1 + 1 > 2r + \alpha$, the value of $\lambda_{j,k}$ is given by formula (7.10). Furthermore, Assumptions (A1)–(A5) hold with $D = D_0 = 0$ if $\alpha < 1$, and with $D = d_U$ and $D_0 = d_T$ if $\alpha \geq 1$. Conditions (5.9)–(5.11) of Lemma 2 are also valid.*

Again, analogously to Section 7.1, due to Proposition 2, all statements and constructions in Sections 3–6 can be applied to equation (1.1) with operator Q given in (7.11). In particular, Theorems 1–3 can be used with $\beta = 2r$.

If q is the Dirac delta function, so that $(Qf)(x) = f(G^{-1}(x))$, then $r = 0$ and the problem reduces to regression estimator based on spatially inhomogeneous data studied in Antoniadis, Pensky and Sapatinas (2012). In this case, the rates of convergence coincide with the minimax convergence rates derived therein.

Remark 6 (Irregularly spaced observations and heterogeneous noise) It follows from examples in Sections 7.1 and 7.2 that there is a direct correspondence between deconvolution with irregularly spaced measurements and deconvolution with heterogeneous noise. In particular, as far as convergence rates are concerned, the squared noise level acts in a similar way to the inverse of the design density and both are equivalent in some way to a multiplicative factor in the convolution operator.

8 Simulation study and real data application

8.1 Simulation study

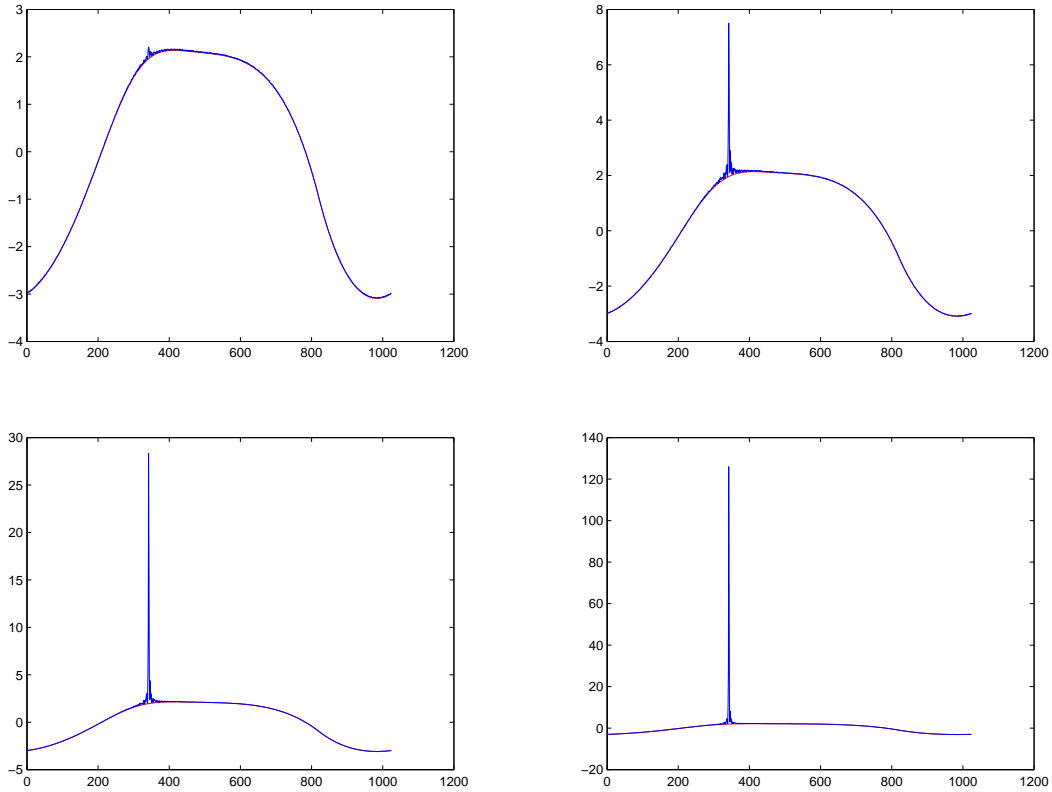


Figure 8.1: True function H (red line) and observed data (blue line) with $\alpha = 1$ (upper left), $\alpha = 2$ (upper right), $\alpha = 2.5$ (lower left) and $\alpha = 3$ (lower right). Here, $\text{SNR} = 0.8848$ for $\alpha = 1$, $\text{SNR} = 0.0808$ for $\alpha = 2$, $\text{SNR} = 0.0183$ for $\alpha = 2.5$ and $\text{SNR} = 0.0040$ for $\alpha = 3$.

In order to assess finite sample properties of the proposed methodology and, in particular, performance of the hybrid estimator, we carried out a small simulation study. We limited our

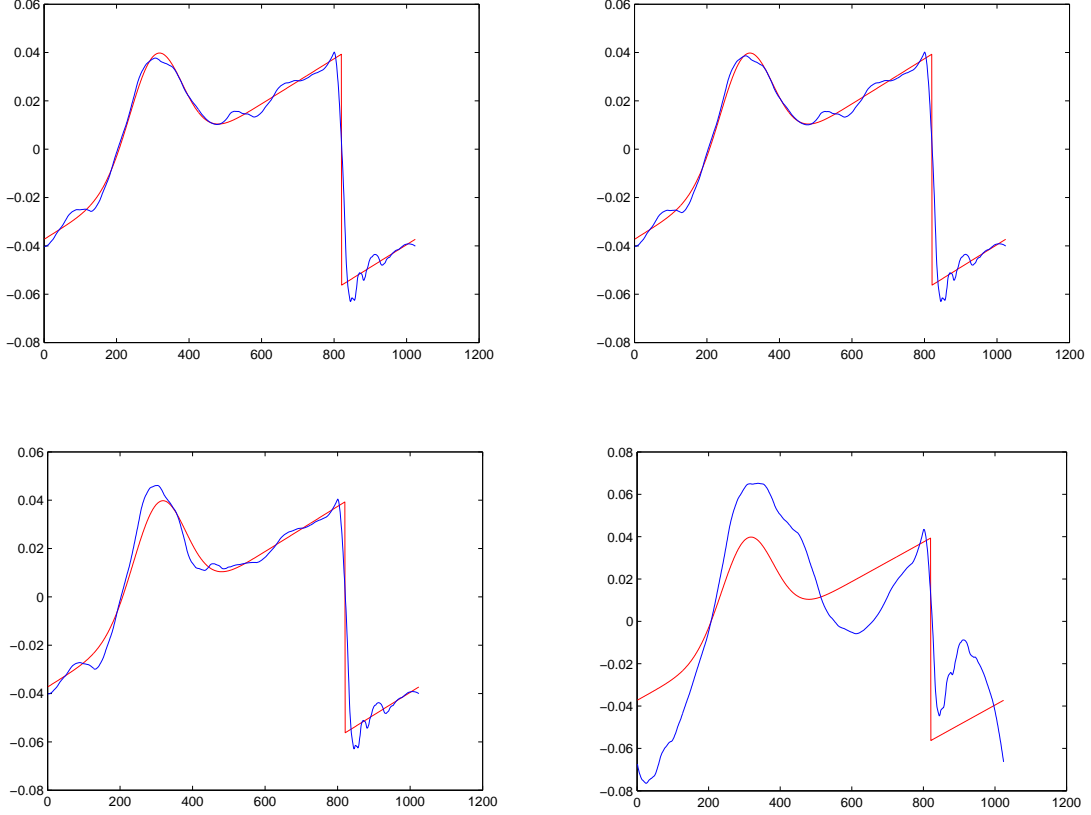


Figure 8.2: Thresholded wavelet-vaguelette deconvolution estimator with $n = 1024$. True regression f (red line) and its estimated value (blue line) with $\alpha = 0$ (upper left), $\alpha = 1$ (upper right), $\alpha = 2$ (lower left) and $\alpha = 3$ (lower right).

attention to the deconvolution in the presence of heteroscedastic noise described in Section 7.1. Specifically, we considered $q(x) = q_1(x)$ with $\lambda = 5$ where $q_1(x)$ is defined in (7.7). We chose one of the standard test functions, **blip**, as the true function $f(x)$. Function $\mu(x)$ in (7.1) is of the form $\mu(x) = |h^{-1}(x - x_0)|^{\alpha/2} \mathbb{I}(|x - x_0| \leq h) + \mathbb{I}(|x - x_0| > h)$ with $x_0 = 1/3$ and $h = 1/6$, so that condition (7.4) holds. We generated data using equation (7.3) with $\gamma(x) = \mu^{-1}(x)$ and $\sigma = 0.02$, in particular,

$$Y_i = H(i/n) + \sigma \gamma(i/n) \xi_i, \quad i = 1, \dots, n, \quad \text{with} \quad H(x) = \int_0^1 q(x-t)f(t)dt.$$

We evaluate noise intensity by the, common in signal processing, signal-to-noise ratio (SNR) which is defined as $\text{SNR} = \sqrt{n} \text{std}(f) / (\|\gamma\| * \sigma)$ where $\|\gamma\|$ is the L^2 -norm of γ and $\text{std}(p)$ is the standard deviation of p for any function $p(x)$.

We used WaveLab package for Matlab and carried out simulations using degree 8 Daubechies wavelets and $n = 1024$. In order to obtain estimators of wavelet and scaling coefficients, we generated wavelet and scaling functions ψ_{jk} and φ_{mk} using **MakeWavelet** command and obtained a respective matrix of the Fourier coefficients. Subsequently, we found estimators of wavelet and scaling coefficients using formula (7.6) with Y_ω being discrete Fourier transform of vector Y in (7.3). We generated values of $\lambda_{j,k}^{-2}$ using equation (7.8) and used them for hard thresholding. Due to relatively small value of n , we did not use block thresholding described in Section 4. By applying inverse

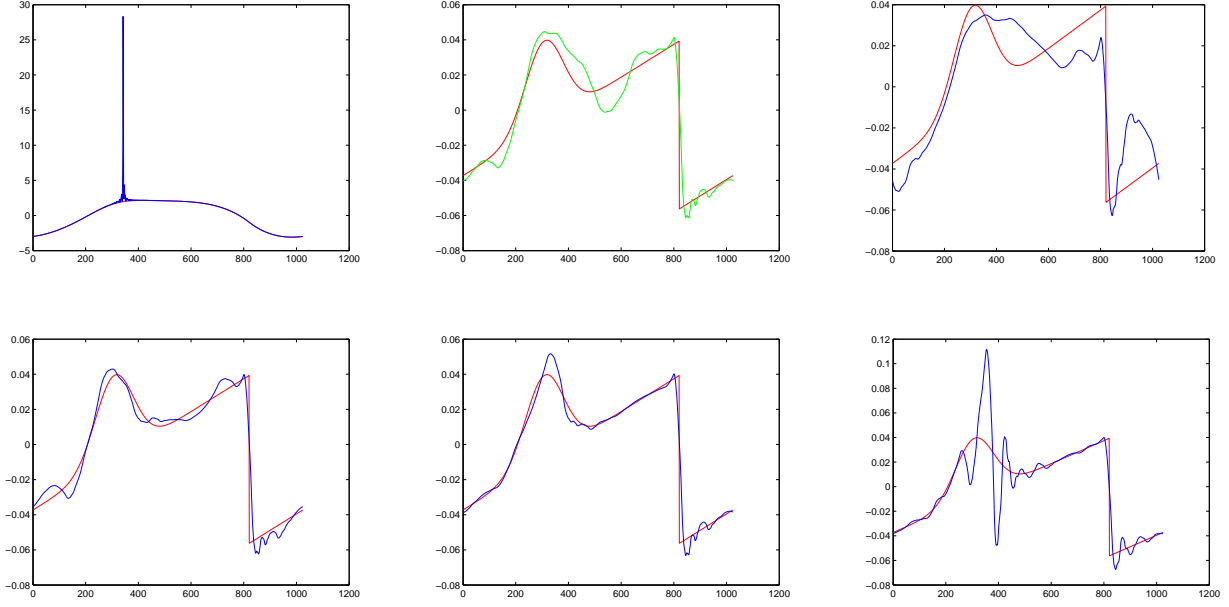


Figure 8.3: Hybrid estimation with $\alpha = 2.5$. True H (red line) and observed data (blue line) (upper left), true f (red line) and wavelet-vaguelette estimator (green line) (upper middle), true f (red line) and hybrid estimators (blue line) with $m = 2$ (upper right), $m = 3$ (lower left), $m = 4$ (lower middle) and $m = 5$ (lower right). Lepskii method selects estimator with $\hat{m} = 4$ (lower middle).

wavelet transform to the thresholded wavelet coefficients, we obtained deconvolution estimator \hat{f} .

We evaluated performance of the estimators for $n = 1024$ and different values of α . As it is expected, when α is growing, the SNRs are decreasing and the quality of observed data is declining. Figure 8.1 demonstrates observed data for various values of α . The corresponding signal-to-noise ratios are $\text{SNR} = 0.8848$ for $\alpha = 1$, $\text{SNR} = 0.0808$ for $\alpha = 2$, $\text{SNR} = 0.0183$ for $\alpha = 2.5$ and $\text{SNR} = 0.0040$ for $\alpha = 3$. Figure 8.2 shows wavelet-vaguelette deconvolution estimators (4.14) obtained for $\alpha = 0$, $\alpha = 1$, $\alpha = 2$ and $\alpha = 3$. As the values of α grow, SNR declines and the wavelet-vaguelette estimators (4.14) deteriorate. If $\alpha = 3$, the wavelet-vaguelette reconstruction has little resemblance to the regression function which it estimates. Note that for moderate values of α , the wavelet-vaguelette estimator adjusts to spatially inhomogeneous noise quite well. Indeed, fluctuations at the right end of the graph appear even when $\alpha = 0$ (upper left) and are due to the relatively crude choice of threshold in formula (4.14). Actually, for $\alpha = 0$ the noise cease to be inhomogeneous and estimator (4.14) reduces to Fourier-wavelet estimator of Johnstone, Kerkyacharian, Picard and Raimondo (2004).

For large values of α , we construct hybrid estimators described in Sections 4 and 6. Construction of the adaptive hybrid estimator consists of the following steps.

1. Fix the lowest resolution level m_1 and the highest resolution level J . For each value of $m = m_1, \dots, J - 1$, repeat steps 2 through 6.
2. Obtain the wavelet-vaguelette estimator of f with the lowest resolution level m using formula (4.14).
3. Identify sets K_{0m} , K_{0m}^c , K_{1j} and K_{1j}^c for $j = m \dots J - 1$. Also, find set Ω_m .

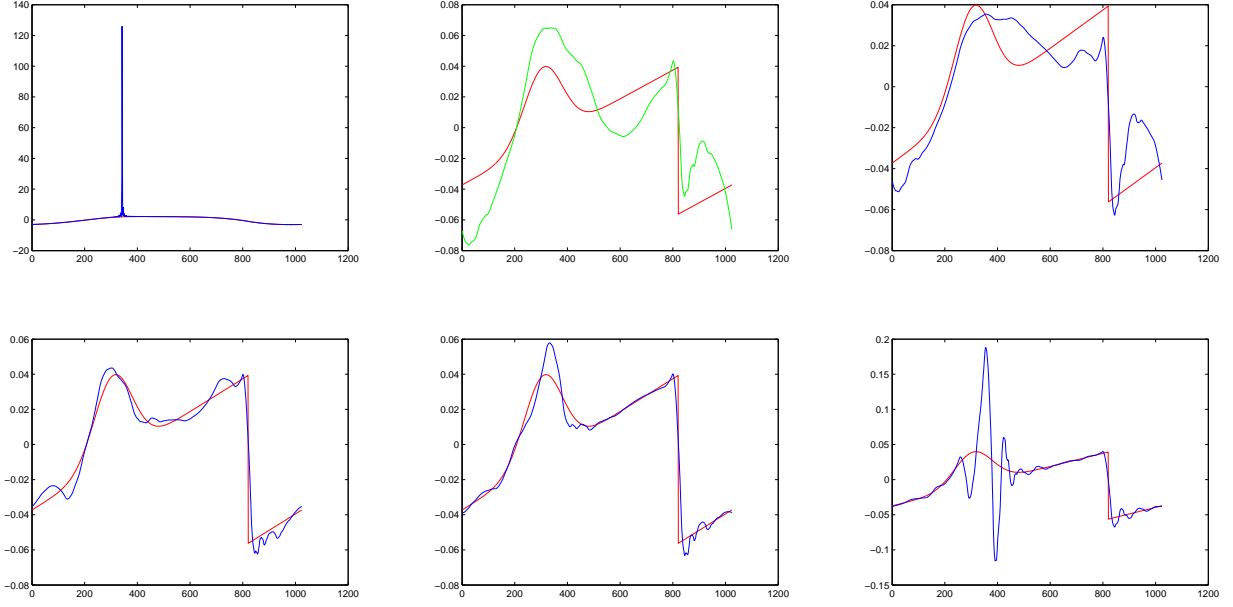


Figure 8.4: Hybrid estimation with $\alpha = 3$. True H (red line) and observed data (blue line) (upper left), true f (red line) and wavelet-vaguelette estimator (green line) (upper middle), true f (red line) and hybrid estimators (blue line) with $m = 2$ (upper right), $m = 3$ (lower left), $m = 4$ (lower middle) and $m = 5$ (lower right). Lepskii method selects estimator with $\hat{m} = 4$ (lower middle).

4. Form matrices $\mathbf{A}^{(m)}$ and $\mathbf{B}^{(m)}$ and vector $\hat{\mathbf{c}}^{(m)}$ using formulae (4.19) and (4.20), respectively, and obtain solution $\hat{\mathbf{z}}^{(m)}$ of the system of equations (4.23). Use vector $\hat{\mathbf{z}}^{(m)}$ as coefficients \hat{a}_{mk} , $k \in K_{0m}$, in the zero-affected portion of the estimator (4.25).
5. Replace estimators of the scaling coefficients (if $k \in K_{0m}$) and wavelet coefficients (if $k \in K_{1j}$, $j = m, \dots, J-1$) by zeros to obtain the zero-free portion of the estimator (4.14).
6. Combine wavelet coefficients in steps 4 and 5 to obtain wavelet coefficients of \hat{f}_m . Recover \hat{f}_m using inverse wavelet transform. Set

$$\lambda_m^{-2} = \sigma^2 \sum_k \mu^{-2}(2^{-m}k) \|T_{mk}\|^2.$$

7. For each $m = m_1, \dots, J-1$, and $j = m+1, \dots, J-1$, evaluate matrix of the adjusted differences

$$\mathcal{L}_{mj} = \|(\hat{f}_m - \hat{f}_j)\mathbb{I}(\Omega_m)\|^2 \Big/ \left(\sigma^2 n^{-1} \log n \lambda_j^{-2} \right). \quad (8.1)$$

Choose \hat{m} in (6.2) as

$$\hat{m} = \min \left\{ m : \mathcal{L}_{mj} \leq \kappa^2 \text{ for all } j, m \leq j \leq J-1 \right\}, \quad (8.2)$$

i.e., by comparing maximum value of row m of matrix \mathcal{L} with a constant κ^2 .

In our simulations, we chose $n = 1024$, $m_1 = 1$ and $J = 7$ and carried out hybrid estimation with $\alpha = 2.5$, $\alpha = 3$ and $\alpha = 4$. Simulation results for these three cases are presented in Figures 8.3, 8.4 and 8.5, respectively. The upper left figures represent the observed noisy data, the upper middle

figures exhibit reconstructions of f by wavelet-vaguelette estimator while the rest of the figures display hybrid estimators of f for resolution levels $m = 2$ to $m = 5$. Observe that for $\alpha = 2.5$ the wavelet-vaguelette estimator still generally follows the true function f but for $\alpha = 3$ or $\alpha = 4$ it bears little resemblance to f . The hybrid estimator allows to account for inhomogeneity of the noise and to significantly improve reconstruction of f .

Lepskii procedure provides a choice of resolution level \hat{m} for each of the values of α . In the case of $\alpha = 2.5$, the row maximum exceeds 93 for $m \leq 3$ and is below 0.2 if $m \geq 4$. In the case of $\alpha = 3$, the row maximum exceeds 20 for $m \leq 3$ and is below 0.3 if $m \geq 4$. In both of these cases, Lepskii method chooses $\hat{m} = 4$. If $\alpha = 4$, the row maximum is very large for $m \leq 2$ and is below 3 for $m \geq 3$, so that $\hat{m} = 3$. Figures 8.3, 8.4 and 8.5 confirm that Lepskii procedure makes correct choices.

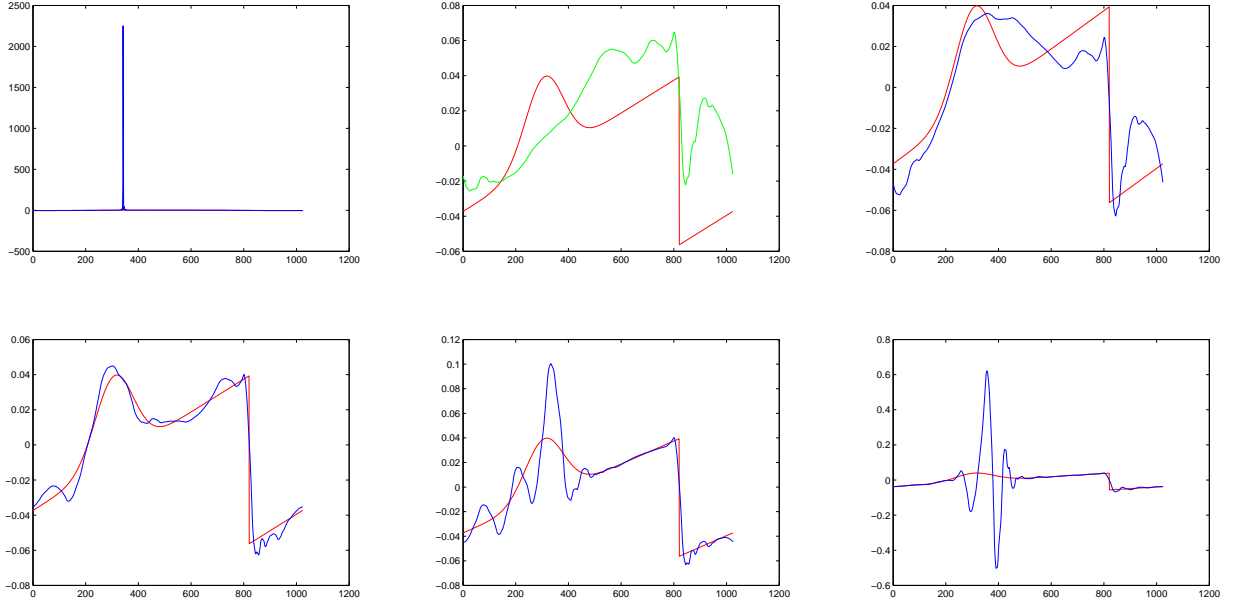


Figure 8.5: Hybrid estimation with $\alpha = 4$. True H (red line) and observed data (blue line) (upper left), true f (red line) and wavelet-vaguelette estimator (green line) (upper middle), true f (red line) and hybrid estimators (blue line) with $m = 2$ (upper right), $m = 3$ (lower left), $m = 4$ (lower middle) and $m = 5$ (lower right). Lepskii method selects estimator with $\hat{m} = 3$ (lower left).

8.2 Real data application

Below, we consider application of the hybrid estimator developed in the paper to recovery of a convolution signal transmitted via Amplitude Modulation described in Example 3. Mathematically, the problem reduces to deconvolution with a spatially inhomogeneous kernel in Section 7.1 and appears in the form of equation (7.2) with

$$\mu(x) = \cos(2\pi\omega x + \theta) \quad (8.3)$$

with $\omega \approx n/2$ and $\theta \in [0; 2\pi]$. We chose $\omega = n/2 + 1$ and expressed θ as $\theta = 3\pi/2 - 2\pi\theta_0$. Then, $\mu(x)$ can be presented as $\mu(x) = \sin(2\pi(n/2 + x - \theta_0))$, so that $\mu(x)$ has two zeros of order $\alpha = 2$, x_{01} and x_{02} , in $[0, 1]$.

For simplicity, we considered the same set up as in simulation example, that is, we used $q(x) = q_1(x)$ with $\lambda = 5$ where $q_1(x)$ is defined in (7.7) and one of the standard test functions, `blip`, as the true function $f(x)$. We carried out simulations with $n = 512$, $\theta_0 = 1/3$ (so that, $x_{01} = 1/3$ and $x_{02} = 5/6$), $\sigma = 0.01$ and degree 8 Daubechies wavelets. The locations of zeros were estimated from the data. One of zeros was placed at the position where the value of the original signal is minimal in absolute value, and another zero was placed $1/2$ units away from the first zero. In our study, locations of zeros were estimated as $\hat{x}_{01} = 0.33496$ and $\hat{x}_{02} = 0.83496$.

The top row of Figure 8.6 presents signal y with uniform noise generated according to equation (7.2) as well as signals Y with heteroscedastic noise obtained according to equation (7.3) by dividing equation (7.2) by $\mu(i/n)$. The bottom row of 8.6 shows the wavelet-vaguelette deconvolution estimator and the hybrid estimators. It is easy to see that the wavelet-vaguelette deconvolution estimator delivers a poor reconstruction while the hybrid technique produces a much more precise estimator of the unknown signal f .

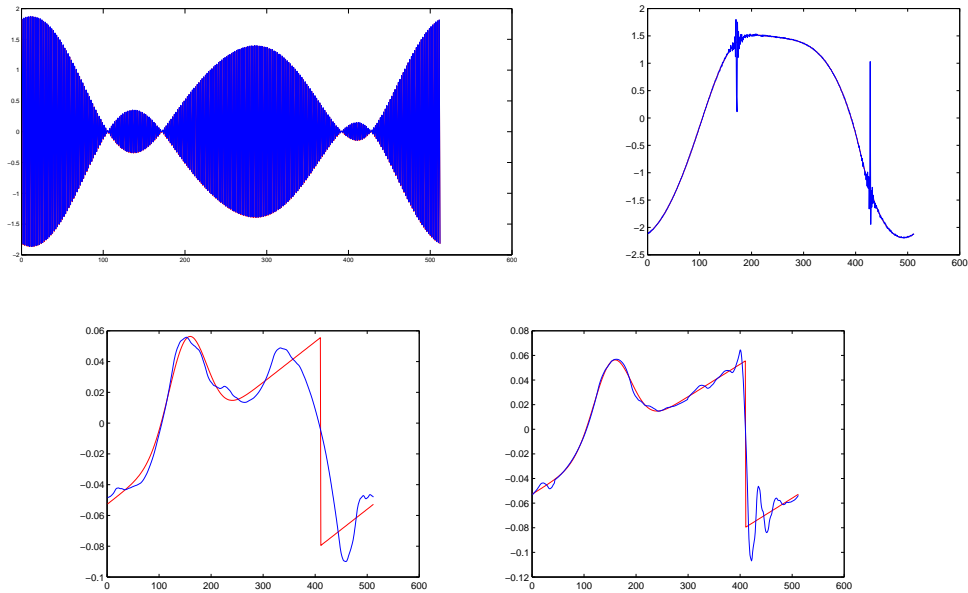


Figure 8.6: Observed values of the signal and estimators of the true function. Top left: true signal (red line) and observed data with homogeneous noise (blue line). Top right: true signal divided by μ (red line) and observed data with heteroscedastic noise (blue line). Bottom left: true function f (red line) and wavelet-vaguelette estimator (blue line). Bottom right: true function f (red line) and hybrid estimator (blue line). Here, $\theta_0 = 1/3$, noise level $\sigma = 0.01$.

9 Discussion

In the present paper, we consider estimation of a solution of a spatially inhomogeneous linear inverse problem (1.1) with possible singularities. The special feature of problems like this is that the degree of ill-posedness depends not only on the scale but also on location. In spite of a huge number of publications devoted to linear inverse problems, to the best of our knowledge, this type of problems has never been treated before. Spatially inhomogeneous ill-posed problems appear naturally when the noise level is location dependent or observations are irregularly spaced. We consider a version of a spatially inhomogeneous problem where there exists a singularity point x_0 such that the norm of the solution grows when the right-hand side is localized in the vicinity of x_0 . This assumption

corresponds to the situation of locally extreme noise and extremely inhomogeneous design. We also assume that the unknown function f belongs to a Besov space and characterize ill-posedness and spatial inhomogeneity of operator Q in terms of wavelet-vaguelette decomposition. The novel feature here is that the norms of vaguelette depend on location and may be infinite in the vicinity of a singularity point, so that SVD-type solutions cease to work.

For this reason, estimators obtained in the paper are based either on wavelet-vaguelette decomposition (if the norms of all vaguelettes are finite) or on a hybrid of wavelet-vaguelette decomposition and Galerkin method (if vaguelettes in the neighborhood of the singularity point have infinite norms). The hybrid estimator is a combination of a linear part in the vicinity of the singularity point and the nonlinear block thresholding wavelet estimator elsewhere. To attain adaptivity, we first choose an optimal resolution level for the linear, singularity affected, portion of the estimator using Lepskii (1990, 1999) method and then use this resolution level as the lowest for nonlinear wavelet estimator. We show that, up to a logarithmic factor, the hybrid estimator attains the asymptotically optimal (in the minimax sense) convergence rates.

The theory presented in the paper is supplemented by examples of deconvolution with a spatially inhomogeneous kernel, deconvolution in the presence of locally extreme noise or extremely inhomogeneous design. The first two problems are examined via a limited simulation study which demonstrates advantages of the hybrid estimator when the degree of spatial inhomogeneity is high. In addition, we apply the technique to recovery of a convolution signal transmitted via amplitude modulation.

We note that the wavelet-based estimation procedure presented in the paper is motivated by the need of constructing an asymptotically optimal estimator in the case when the unknown function f is spatially inhomogeneous. The estimator uses relatively crude thresholding procedure which can be improved by applying more sensitive thresholding techniques. Moreover, one can possibly find more efficient computational procedures than the hybrid estimator if establishing asymptotic optimality is not a priority.

Finally, in the paper, we consider only the simplest case when the unknown function is univariate and is defined on an interval. The problem can be naturally extended to the case of multivariate function f which belongs to an isotropic or anisotropic Besov space. The paper assumes that the operator Q in (1.1) is completely known. However, if this is not true, it will be interesting to investigate how uncertainty about Q affects the rates of convergence. Also, although the hybrid estimator works adequately when Q is completely known, it would require appropriate modifications if Q is partially unknown. However, all these extensions will be a matter of future investigation.

Acknowledgements

Marianna Pensky was supported in part by National Science Foundation (NSF), grant DMS-1106564. We would like to thank an Associate Editor and two anonymous referees for their suggestions which helped to significantly improve the paper.

10 Proofs

10.1 Lower bounds

Proof of Theorem 1. The rates are derived by standard methods described in, e.g., Tsybakov (2009). For this reason, we shall provide a very brief proof.

The main idea of the proof is based on Lemma A.1 of Bunea, Tsybakov and Wegkamp (2007). In order to show that, for some C and $H > 0$,

$$R_\epsilon(B_{p,q}^s(A)) \geq CH, \quad (10.1)$$

one needs to find a subset of functions $\mathcal{F} \subset B_{p,q}^s(A)$ such that for any pair $f_1, f_2 \in \mathcal{F}$,

$$\|f_1 - f_2\|^2 \geq 4H \quad (10.2)$$

and the Kullback-Leibler divergence

$$\mathbb{K}(\mathbb{P}_{f_1}, \mathbb{P}_{f_2}) = 0.5 \varepsilon^{-1} \|\mathbf{q}_1 - \mathbf{q}_2\|^2 / 2 \leq \ln \text{card}(\mathcal{F}) / 16. \quad (10.3)$$

We consider two cases here: the strongly inhomogeneous and the weakly inhomogeneous cases. In the strongly inhomogeneous case, the hardest set to estimate is the finite set of functions which are concentrated around singularity point. In the weakly inhomogeneous case, this set is comprised of functions which are uniformly distributed over some resolution level.

The strongly inhomogeneous case Consider a set of functions $\mathcal{F} = \{\gamma_j \psi_{jk} : |k - k_{0j}| \leq K\}$ where K is a fixed positive constant. Then, $\text{card}(\mathcal{F}) \geq 2K - 1$. In order $f \in B_{p,q}^s(A)$, one needs $\gamma_j \leq A2^{-js'}$, so set $\gamma_j = A2^{-js'}$. It is easy to check that for $f_i = \gamma_j \psi_{jk_i}$, $i = 1, 2$, one has

$$\|\mathbf{q}_1 - \mathbf{q}_2\|^2 = \gamma_j^2 \|v_{jk_1} - v_{jk_2}\|^2 \leq C \gamma_j^2 (\lambda_{jk_1}^2 + \lambda_{jk_2}^2) \asymp 2^{-j(\alpha+\beta)} \gamma_j^2,$$

by (2.6), since $|k - k_{0j}| \leq K$. Hence, it follows from (10.3) that j is such that $2^j \asymp (A^2/\varepsilon)^{1/(2s'+\alpha+\beta)}$. Note that $\|f_1 - f_2\|^2 = 2\gamma_j^2$, so that $H = 0.5 A^2 2^{-2js'}$ and, therefore,

$$R_\varepsilon(B_{p,q}^s(A)) \geq C A^{\frac{2(\alpha+\beta)}{2s'+\alpha+\beta}} \varepsilon^{\frac{2s'}{2s'+\alpha+\beta}}. \quad (10.4)$$

The weakly inhomogeneous case Consider a set Ω of binary sequences of length $N = 2^j$, $j \geq 3$:

$$\Omega^* = \{\omega = (\omega_0, \dots, \omega_{N-1}), \omega_i \in \{0, 1\}\} = \{0, 1\}^N$$

and a corresponding set of functions

$$\mathcal{F}^* = \left\{ f_\omega = \gamma_j \sum_{k=0}^{N-1} \omega_k \psi_{jk}, \omega \in \Omega \right\}.$$

Let $\rho(\omega, \omega') = \sum_k \mathbb{I}(\omega_k \neq \omega'_k)$. Then, for any $\omega, \omega' \in \Omega$, one has $\|f_\omega - f_{\omega'}\|^2 = \gamma_j^2 \rho(\omega, \omega')$. By Varshamov-Gilbert Lemma (see Lemma 2.9 of Tsybakov (2009)), there exists a subset $\Omega = (\omega^{(0)}, \omega^{(1)}, \dots, \omega^{(M)})$ of Ω^* such that $M \geq 2^{N/8}$ and $\rho(\omega, \omega') \geq N/8$ for any $\omega, \omega' \in \Omega$.

Consider subset $\mathcal{F} = \{f_\omega \in \mathcal{F}^* : \omega \in \Omega\}$. Then, for any $f_\omega, f_{\omega'} \in \mathcal{F}$, one has

$$\|f_\omega - f_{\omega'}\|^2 \geq 2^j \gamma_j^2 / 64, \quad \text{card}(\mathcal{F}) \geq \ln 2 \cdot 2^j / 8. \quad (10.5)$$

Since $f_\omega \in B_{p,q}^s(A)$, we set $\gamma_j = A2^{-j(s+1/2)}$. Now it remains to determine relationship between j and ε . For this purpose, note that, by (2.3) and (2.6),

$$\begin{aligned} \|Qf_\omega - Qf_{\omega'}\|^2 &= \gamma_j^2 \left\| \sum_{k=0}^{2^j-1} \mathbb{I}(\omega_k \neq \omega'_k) v_{j,k} \right\|^2 \leq \gamma_j^2 \sum_{k=0}^{2^j-1} \lambda_{j,k}^2 \\ &\leq C 2^{j(\alpha+\beta)} 2^{j(\alpha+1)} \gamma_j^2. \end{aligned}$$

Hence, relations (10.3) and the second inequality in (10.5) imply that $2^j \asymp (A^2/\varepsilon)^{1/(2s+\beta+1)}$. Therefore, by (10.1) and the first inequality in (10.5), one has

$$R_\varepsilon(B_{p,q}^s(A)) \geq C A^{\frac{2(\beta+1)}{2s+\beta+1}} \varepsilon^{\frac{2s}{2s+\beta+1}}. \quad (10.6)$$

Now, to complete the proof of Theorem 1, note that the lower bound given by formula (10.6) dominates the one given by (10.4) if $2s(\alpha - 1) \geq (\beta + 1)(1 - 2/p)$ and visa versa.

10.2 Supplementary large deviation results

Lemma 4 *Let $\tau_0^2 = C_u C_\lambda (\sqrt{2}\chi + 1)^2$. Then*

$$\mathbb{P} \left(\sum_{k \in U_{jl}} (\hat{b}_{jk} - b_{jk})^2 > \tau_0^2 R_{jl\varepsilon} \right) \leq \varepsilon^\chi, \quad (10.7)$$

where $R_{jl\varepsilon}$ is defined in (4.13).

Proof of Lemma 4. Consider the set of vectors

$$\Omega_{jl} = \left\{ \nu_k, k \in U_{jl} : \sum_{k \in U_{jl}} |\nu_k|^2 \leq 1 \right\},$$

and the centered Gaussian process defined by

$$Z_{jl}(\nu) = \sum_{k \in U_{jl}} \nu_k (\hat{b}_{jk} - b_{jk}).$$

The proof of the lemma is based on the following inequality:

Lemma 5 (Cirelson, Ibragimov & Sudakov (1976)). *Let D be a subset of $\mathbb{R} = (-\infty, \infty)$, and let $(\xi_t)_{t \in D}$ be a centered Gaussian process. If $\mathbb{E}(\sup_{t \in D} \xi_t) \leq B_1$ and $\sup_{t \in D} \text{Var}(\xi_t) \leq B_2$, then, for all $x > 0$, we have*

$$\mathbb{P} \left(\sup_{t \in D} \xi_t \geq x + B_1 \right) \leq \exp \left(-x^2 / (2B_2) \right). \quad (10.8)$$

To apply Lemma 5, note that

$$\sup_{\nu \in \Omega_{jl}} Z_{jl}(\nu) = \left[\sum_{k \in U_{jl}} |\hat{b}_{jk} - b_{jk}|^2 \right]^{1/2}.$$

Hence, by Jensen's inequality, we derive that

$$B_1 \leq \left[C_u \varepsilon \sum_{k \in U_{jl}} \lambda_{j,k}^{-2} \right]^{1/2} \leq \sqrt{C_u C_\lambda R_{jl\varepsilon}}.$$

Also, by assumption (A2),

$$B_2 = \sup_{\nu \in \Omega_{jl}} \text{Var}(Z_{jl}(\nu)) = \varepsilon \sup_{\nu \in \Omega_{jl}} \left\| \sum_{k \in U_{jl}} u_{j,k} \nu_k \right\|^2 \leq \varepsilon C_u \max_{k \in U_{jl}} \lambda_{j,k}^{-2} \leq C_u C_\lambda R_{jl\varepsilon}.$$

Therefore, by applying Lemma 5 with $x^2 = 2\chi^2 C_u C_\lambda R_{jl\varepsilon}$ and noting that, therefore, $x + B_1 = \tau_0$, we obtain (10.7).

Lemma 6 Let m_0 and \hat{m} be given by formulae (5.14) and (6.2), respectively, and $\hat{\mathbf{z}}^{(m)}$ be the solution of the system of equations given by (4.24). Let assumptions of Lemma 2 hold. Denote $\eta_{m\varepsilon} = \lambda_m^{-1} \sqrt{\varepsilon \ln(\varepsilon^{-1})}$. If $m > m_0$, then, for any $\nu > 0$, as $\varepsilon \rightarrow 0$,

$$\mathbb{P}\left(\|\hat{\mathbf{z}}^{(m)} - \mathbf{z}^{(m)}\| > \nu \eta_{m\varepsilon}\right) \leq C\varepsilon^{\nu^2/C_\nu}, \quad (10.9)$$

where

$$C_\nu = 32D_0^2 C_{A_2}^2 \max(C_{\rho\lambda}^{-1} C_{w2}^2, C_t K_1) \quad (10.10)$$

and constants C_{A_2} , $C_{\rho\lambda}$, C_{w2} , C_t and K_1 are defined in (5.8), (5.9), (5.7), (2.4) and (5.10), respectively.

Proof of Lemma 6. Observe that for any m , by (4.24), one has

$$\|\hat{\mathbf{z}}^{(m)} - \mathbf{z}^{(m)}\| \leq \|(\mathbf{A}^{(m)})^{-1}(\hat{\mathbf{c}}^{(m)} - \mathbf{c}^{(m)})\| + \|(\mathbf{A}^{(m)})^{-1}\mathbf{B}^{(m)}(\hat{\mathbf{h}}^{(m)} - \mathbf{h}^{(m)})\| + \|(\mathbf{A}^{(m)})^{-1}\mathbf{r}^{(m)}\|.$$

Recall that, by (10.40) and (10.43), inequality $\|(\mathbf{A}^{(m)})^{-1}\mathbf{r}^{(m)}\|^2 \leq 2DC_{A_2}^2 A^2 K_2 2^{-2ms'}$ holds, and observe that, for $m > m_0$, $\|(\mathbf{A}^{(m)})^{-1}\mathbf{r}^{(m)}\| = o(\eta_{m\varepsilon})$ as $\varepsilon \rightarrow 0$. Therefore, as $\varepsilon \rightarrow 0$,

$$\begin{aligned} \mathbb{P}\left(\|\hat{\mathbf{z}}^{(m)} - \mathbf{z}^{(m)}\| > \nu \eta_{m\varepsilon}\right) &\leq \mathbb{P}\left(\|(\mathbf{A}^{(m)})^{-1}(\hat{\mathbf{c}}^{(m)} - \mathbf{c}^{(m)})\| > 0.5\nu \eta_{m\varepsilon}(1 - o(1))\right) \\ &+ \mathbb{P}\left(\|(\mathbf{A}^{(m)})^{-1}\mathbf{B}^{(m)}(\hat{\mathbf{h}}^{(m)} - \mathbf{h}^{(m)})\| > 0.5\nu \eta_{m\varepsilon}(1 - o(1))\right) \equiv P_1 + P_2. \end{aligned}$$

Here, since $\hat{\mathbf{c}}^{(m)} - \mathbf{c}^{(m)}$ is a $(2D_0)$ -dimensional normal vector with zero mean and the component variances $\text{Var}(\hat{c}_l^{(m)}) = \varepsilon \|w_{m,l}\|^2$, using formulae (5.7), (5.9) and (5.8), one derives, as $\varepsilon \rightarrow 0$,

$$\begin{aligned} P_1 &\leq \mathbb{P}\left(\|\hat{\mathbf{c}}^{(m)} - \mathbf{c}^{(m)}\| > 0.5C_{A_2}^{-1}\nu \rho_m^2 \eta_{m\varepsilon}(1 - o(1))\right) \\ &\leq 2D_0 \max_{l \in K_{0m}} \mathbb{P}\left(|\hat{c}_l^{(m)} - c_l^{(m)}| > (4C_{A_2}D_0)^{-1} \sqrt{C_{\rho\lambda} \varepsilon \ln(\varepsilon^{-1})} \nu \rho_m(1 - o(1))\right) \leq 2D_0 \varepsilon^{\nu^2/C_{\nu_1}}, \end{aligned} \quad (10.11)$$

where $C_{\nu_1} = 32D_0^2 C_{A_2}^2 C_{w2}^2 C_{\rho\lambda}^{-1}$.

For the P_2 term, note that $\xi = \mathbf{B}^{(m)}(\hat{\mathbf{h}}^{(m)} - \mathbf{h}^{(m)})$ is a $(2D_0)$ -dimensional normal vector with zero mean and the component variances

$$\text{Var}(\xi_l) = \varepsilon \left\| \sum_{k \in K_{0m}^c} t_{m,k} \langle w_{m,k}, w_{m,l} \rangle \right\|^2 \leq \varepsilon C_t K_1 \lambda_m^{-2} \rho_m^4$$

by Assumption (A4) and inequality (5.10). Then, using (5.8), similarly to the case of P_1 , one obtains

$$\begin{aligned} P_2 &\leq \mathbb{P}\left(\|\mathbf{B}^{(m)}(\hat{\mathbf{h}}^{(m)} - \mathbf{h}^{(m)})\| > 0.5C_{A_2}^{-1}\nu \rho_m^2 \eta_{m\varepsilon}(1 - o(1))\right) \\ &\leq 2D_0 \max_{l \in K_{0m}} \mathbb{P}\left(|\xi_l| > \nu (4C_{A_2}D_0)^{-1} \rho_m^2 \eta_{m\varepsilon}(1 - o(1))\right) \\ &\leq 2D_0 \max_{l \in K_{0m}} \mathbb{P}\left(|\xi_l|/\sqrt{\text{Var}\xi_l} > \nu (4C_{A_2}D_0)^{-1} \sqrt{(C_t K_1)^{-1} \ln(\varepsilon^{-1})}\right) \leq 2D_0 \varepsilon^{\nu^2/C_{\nu_2}}, \end{aligned} \quad (10.12)$$

where $C_{\nu_2} = 32D_0^2 C_{A_2}^2 C_t K_1$. Combination of (10.11) and (10.12) completes the proof of the lemma.

10.3 Proofs of statements in Section 5

Proof of Lemma 1 is based on the following statements.

Lemma 7 *Let $1 \leq p, q \leq \infty$, $s \geq \max(1/2, 1/p)$, and Assumptions (A1)–(A4) hold. Let B_{jl} and $R_{jl\varepsilon}$ be defined in (4.12) and (4.13), respectively. Then,*

$$\sup_{f \in B_{p,q}^s(A)} \sum_{j=m}^{J-1} \sum_{l \in U_j} \min(B_{jl}, R_{jl\varepsilon}) \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho, \quad (10.13)$$

where $\Delta(\varepsilon)$ and ρ are defined in (3.3) and (5.5), respectively.

Proof of Lemma 7. First, note that

$$\sum_{l \in U_j} R_{jl\varepsilon} = R_{j\varepsilon}^* \leq \begin{cases} C_\alpha [l_{j\varepsilon}]^{(1-\alpha)_+}, & \text{if } \alpha \neq 1, \\ C_\alpha \ln(l_{j\varepsilon}), & \text{if } \alpha = 1. \end{cases} \quad (10.14)$$

Here, $x_+ = \max(x, 0)$, $l_{j\varepsilon}$ is defined in formula (4.8) and constant C_α depends on α only. Recall also that if $f \in B_{p,q}^s(A)$, then

$$\sum_{k=0}^{2^j-1} b_{jk}^2 \leq C^* 2^{-2js^*}, \quad |b_{jk}| \leq A 2^{-js'}. \quad (10.15)$$

Note that it follows from (10.14) and (10.15) that

$$\begin{aligned} D_1(j_1) &= \sum_{j=m}^{j_1} \sum_{l \in U_j} R_{jl\varepsilon} \leq 2C_\alpha 2^{j_1(\beta + \max(1, \alpha))} \varepsilon [\ln(\varepsilon^{-1})]^{\mathbb{I}(\alpha=1) - (\alpha-1)_+}, \\ D_2(j_2) &= \sum_{j=j_2}^{J-1} \sum_{l \in U_j} B_{jl} \leq C^* 2^{-2j_2 s^*}. \end{aligned}$$

We consider the strongly inhomogeneous and the weakly inhomogeneous cases separately.

The strongly inhomogeneous case. Let $2s(\alpha - 1) > (\beta + 1)(1 - 2/p)$. Choose j_i so that

$$D_i(j_i) \leq C \varepsilon^{\frac{2s'}{2s' + \alpha + \beta}}, \quad i = 1, 2. \quad (10.16)$$

It is easy to see that (10.16) holds if j_1 and j_2 are such that

$$2^{j_1} = \varepsilon^{-\frac{\alpha + \beta}{(\max(1, \alpha) + \beta)(2s' + \alpha + \beta)}} [\ln(\varepsilon^{-1})]^{-\frac{\mathbb{I}(\alpha=1)}{2s' + \alpha + \beta}}, \quad 2^{j_2} = \varepsilon^{-\frac{s'}{s^*(2s' + \alpha + \beta)}}. \quad (10.17)$$

Now, we need to evaluate

$$D_3(j_1, j_2) = \sum_{j=j_1}^{j_2-1} \sum_{l \in U_j} \min(B_{jl}, R_{jl\varepsilon}). \quad (10.18)$$

Consider cases of $\alpha > 1$, $\alpha = 1$ and $\alpha < 1$ separately.

If $\alpha > 1$, then $p > 2$ since, otherwise, $s^* = s'$ and $D_3(j_1, j_2) = 0$. Observe that

$$\begin{aligned} D_3(j_1, j_2) &\leq \sum_{j=j_1+1}^{j_2-1} \left[\sum_{|l| \leq N_j} \sum_{k \in U_{jl}} b_{jk}^2 + \sum_{|l| > N_j} R_{jl\varepsilon} \right] \\ &\leq \sum_{j=j_1+1}^{j_2-1} \left[\left(\sum_{|k-k_{0j}| \leq N_j \ln(\varepsilon^{-1})} |b_{jk}|^p \right)^{2/p} [N_j \ln(\varepsilon^{-1})]^{1-2/p} + C \varepsilon 2^{j(\alpha+\beta)} [N_j \ln(\varepsilon^{-1})]^{1-\alpha} \right]. \end{aligned} \quad (10.19)$$

The two terms are of equal order if

$$N_j = C [\ln(\varepsilon^{-1})]^{-1} \left[\varepsilon 2^{j(2s'+\alpha+\beta)} \right]^{p/(p\alpha-2)} \quad (10.20)$$

By direct calculations, one can check that

$$D_3(j_1, j_2) \leq C \sum_{j=j_1+1}^{j_2-1} 2^{j\gamma} \varepsilon^{\frac{p-2}{p\alpha-2}} \quad (10.21)$$

where

$$\gamma = \frac{2s' + \beta(1 - 2/p) - 2\alpha s}{\alpha - 2/p}. \quad (10.22)$$

Note that $\gamma < 0$ in (10.22) since $\alpha > 1 > 2/p$. Therefore, $D_3(j_1, j_2) \leq C 2^{j_1\gamma_1} \varepsilon^{\frac{p-2}{p\alpha-2}} = C \varepsilon^{\frac{2s'}{2s'+\alpha+\beta}}$ and $\rho = 0$ in (10.13).

If $\alpha = 1$, then $p < 2$, so that $s^* = s'$ and 2^{j_1} and 2^{j_2} differ by logarithmic factor only. Then, $D_3(j_1, j_2) \leq C 2^{-2j_1 s'}$ and (10.13) holds with ρ of the form (5.5).

If $\alpha < 1$, then $p < 2$. Note that, in this case,

$$B_{jl}^{p/2} = \left(\sum_{k \in U_{jl}} b_{jk}^2 \right)^{p/2} \leq \sum_{k \in U_{jl}} |b_{jk}|^p,$$

and, therefore,

$$\begin{aligned} D_3(j_1, j_2) &\leq \sum_{j=j_1+1}^{j_2-1} \left[\sum_{|l| \leq N_j} \sum_{k \in U_{jl}} R_{jl\varepsilon} + \sum_{|l| > N_j} B_{jl}^{p/2} R_{jl\varepsilon}^{1-p/2} \right] \\ &\leq C \sum_{j=j_1+1}^{j_2-1} \left[\varepsilon [N_j \ln(\varepsilon^{-1})]^{1-\alpha} 2^{j(\alpha+\beta)} + \left(\varepsilon [\ln(\varepsilon^{-1})]^{1-\alpha} N_j^{-\alpha} 2^{j(\alpha+\beta)} \right)^{1-p/2} 2^{-js'p} \right]. \end{aligned} \quad (10.23)$$

The terms in the sum above are of equal order if

$$N_j = C \left[\varepsilon [\ln(\varepsilon^{-1})]^{1-\alpha} 2^{j(2s'+\alpha+\beta)} \right]^{p/(p\alpha-2)}. \quad (10.24)$$

Then,

$$D_3(j_1, j_2) \leq C \sum_{j=j_1+1}^{j_2-1} 2^{j\gamma} \varepsilon^{\frac{p-2}{p\alpha-2}} [\ln(\varepsilon^{-1})]^{\frac{(1-\alpha)(2/p-1)}{2/p-\alpha}} \quad (10.25)$$

where γ is given by equation (10.22) and is positive. Hence, $\rho = (2/p - \alpha)^{-1}(1 - \alpha)(2/p - 1)$ in (10.13).

The weakly inhomogeneous case. Let $2s(\alpha - 1) < (\beta + 1)(1 - 2/p)$. Choose j_i so that

$$D_i(j_i) \leq C \varepsilon^{\frac{2s}{2s+\beta+1}}, \quad i = 1, 2. \quad (10.26)$$

It is easy to see that (10.26) holds if j_1 and j_2 are such that

$$2^{j_1} = \varepsilon^{-\frac{1+\beta}{(\max(1,\alpha)+\beta)(2s+\beta+1)}} [\ln(\varepsilon^{-1})]^{-\frac{1(\alpha=1)}{2s+\beta+1}}, \quad 2^{j_2} = \varepsilon^{-\frac{s}{s^*(2s+1+\beta)}}. \quad (10.27)$$

Again, consider cases of $\alpha > 1$, $\alpha = 1$ and $\alpha < 1$ separately.

If $\alpha > 1$, then $p \geq 2$ and $s^* = s$. Then, we obtain that $D_3(j_1, j_2)$ is of the form (10.19) and the terms are equal if N_j is given by (10.20). Plugging in the value of N_j , we derive (10.21) with $\gamma > 0$ in (10.22). After that, it is easy to check that $\rho = 0$ in (5.5) and (10.13).

If $\alpha = 1$, then $p \geq 2$, so that $s^* = s'$ and 2^{j_1} and 2^{j_2} again differ by log factor only. Then, $D_3(j_1, j_2) \leq C2^{-2j_1 s'}$ and (3.3) holds with ρ of the form (5.5).

Let $\alpha < 1$. If $p \geq 2$, then $j_1 = j_2$ and $D_3(j_1, j_2) = 0$. If $p < 2$, then $D_3(j_1, j_2)$ is given by (10.23), and the terms in the sum are equal to each other if N_j is of the form (10.24). Then, $D_3(j_1, j_2)$ is of the form (10.25) with $\gamma < 0$. Plugging $2^{\gamma j_1}$ into (10.25) yields (10.13) with $\rho = (2/p - \alpha)^{-1}(1 - \alpha)(2/p - 1)$.

The moderately inhomogeneous case. If $2s(\alpha - 1) = (\beta + 1)(1 - 2/p)$, then $\gamma = 0$ in (10.22). The value of $D_3(j_1, j_2)$ is given by (10.21) or (10.25), and, since $j_2 - j_1 \leq C \ln(\varepsilon^{-1})$ and

$$\frac{p-2}{\alpha p - 2} = \frac{2s'}{2s' + \alpha + \beta} = \frac{2s}{2s + \beta + 1},$$

equation (10.13) holds with ρ of the form (5.5). This completes the proof.

Proof of Lemma 1. Note that

$$\Delta = \mathbb{E} \|\hat{f}_{c,m} - f_c\|^2 = \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4, \quad (10.28)$$

where

$$\begin{aligned} \Delta_1 &= \sum_{k \in K_{0m}^c} \text{Var}(\hat{a}_{mk}), \quad \Delta_2 = \sum_{j=J}^{\infty} \sum_{k \in K_{1j}^c} b_{jk}^2, \\ \Delta_3 &= \sum_{j=m}^{J-1} \sum_{l \in U_j} \left[\sum_{k \in U_{jl}} \mathbb{E}(\hat{b}_{jk} - b_{jk})^2 \mathbb{I}(\hat{B}_{jl} \geq \tau^2 R_{jl\varepsilon}) \right], \\ \Delta_4 &= \sum_{j=m}^{J-1} \sum_{l \in U_j} B_{jl}^2 \mathbb{P}(\hat{B}_{jl} \geq \tau^2 R_{jl\varepsilon}) \end{aligned}$$

with τ defined in (5.3).

By Assumption (A4) and formula (5.1), one has

$$\Delta_1 \leq \varepsilon \sum_{k \in K_{0m}^c} \|t_{m,k}\|^2 \leq \varepsilon C_t^2 \lambda_m^{-2}. \quad (10.29)$$

By (10.15), one has $\Delta_2 \leq C2^{-2Js^*}$ where J is defined in (4.15). Note that if $2s(\alpha - 1) \geq (\beta + 1)(1 - 2/p)$, then j_2 in (10.17) is such that $j_2 \leq J$ since $2s' \geq 1$ and $s^*/s' \leq 1/2$ due to $s \geq \max(1/p, 1/2)$. If $2s(\alpha - 1) < (\beta + 1)(1 - 2/p)$, then j_2 is given by formula (10.27). If $1 \leq p \leq 2$, then $\alpha \leq 1$, $s^* = s'$ and $j_2 \leq J$ since $2s' + s'(1 + \beta)/s \geq 1 + (\alpha + \beta)/2$. If $p > 2$, then $s^* = s$. Moreover, due to $2s\alpha < 2s' + \beta(1 - 2/p)$, one has $\alpha < s'/s + \beta(1/2 - 1/p)/s < (s + 1/2)/s + \beta/(2s) \leq 2 + \beta$ since $s > 1/2$. Then, $1 + (\alpha + \beta)/2 < 1 + 2s + 1 + \beta$ and $j_2 < J$. Hence,

$$\Delta_2 \leq C \Delta(\varepsilon) \quad (10.30)$$

where $\Delta(\varepsilon)$ is defined in formula (3.3).

In order to obtain an upper bound for Δ_3 and Δ_4 , note that

$$\Delta_3 \leq \Delta_{31} + \Delta_{32}, \quad \Delta_4 \leq \Delta_{41} + \Delta_{42}, \quad (10.31)$$

where

$$\begin{aligned}
\Delta_{31} &= \sum_{j=m}^{J-1} \sum_{l \in U_j} \left[\sum_{k \in U_{jl}} \mathbb{E}(\hat{b}_{jk} - b_{jk})^2 \mathbb{I} \left(\sum_{k \in U_{jl}} (\hat{b}_{jk} - b_{jk})^2 > 0.25 \tau^2 R_{jl\varepsilon} \right) \right], \\
\Delta_{32} &= \sum_{j=m}^{J-1} \sum_{l \in U_j} \left[\sum_{k \in U_{jl}} \mathbb{E}(\hat{b}_{jk} - b_{jk})^2 \mathbb{I}(B_{jl} \geq 0.25 \tau^2 R_{jl\varepsilon}) \right], \\
\Delta_{41} &= \sum_{j=m}^{J-1} \sum_{l \in U_j} \left[B_{jl} \mathbb{P} \left(\sum_{k \in U_{jl}} (\hat{b}_{jk} - b_{jk})^2 > 0.25 \tau^2 R_{jl\varepsilon} \right) \right], \\
\Delta_{42} &= \sum_{j=m}^{J-1} \sum_{l \in U_j} [B_{jl} \mathbb{I}(B_{jl} \leq 1.25 \tau^2 R_{jl\varepsilon})].
\end{aligned} \tag{10.32}$$

By Lemma 4 with $\tau_0 = \tau/2$, we obtain

$$\mathbb{P} \left(\sum_{k \in U_{jl}} (\hat{b}_{jk} - b_{jk})^2 > 0.25 \tau^2 R_{jl\varepsilon} \right) \leq \varepsilon^{\chi^2}. \tag{10.33}$$

Hence, since $j \leq J-1$, derive

$$\begin{aligned}
\Delta_{31} &\leq \sum_{j=m}^{J-1} \sum_{l \in U_j} \left[\left(\sum_{k \in U_{jl}} \mathbb{E}(\hat{b}_{jk} - b_{jk})^4 \right)^{1/2} \left(\mathbb{P} \left(\sum_{k \in U_{jl}} (\hat{b}_{jk} - b_{jk})^2 > 0.25 \tau^2 R_{jl\varepsilon} \right) \right)^{1/2} \right] \\
&\leq C \varepsilon^{1+0.5\chi^2} \sum_{j=m}^{J-1} \sum_{k \in K_{1j}^c} \lambda_{j,k}^{-2} \leq C \varepsilon^{1+\frac{\chi^2}{2} - \frac{\beta + \max(1, \alpha)}{\beta + 1 + \min(1, \alpha)}} = O(\varepsilon)
\end{aligned} \tag{10.34}$$

due to assumption (5.3), and since $\mathbb{E}(\hat{b}_{jk} - b_{jk})^4 \leq 3\varepsilon^2 C_u \lambda_{j,k}^{-4}$. In a similar manner, since $b_{jk} = o(1)$, as $\varepsilon \rightarrow 0$, one has

$$\Delta_{41} \leq \sum_{j=m}^{J-1} \sum_{k \in K_{1j}^c} b_{jk}^2 \varepsilon^{\chi^2} = o(\varepsilon). \tag{10.35}$$

In order to find an upper bound for $\Delta_{32} + \Delta_{42}$, note that $\sum_{k \in U_{jl}} \mathbb{E}(\hat{b}_{jk} - b_{jk})^2 \leq C_u C_\lambda R_{jl\varepsilon}$, so that

$$\begin{aligned}
\Delta_{32} + \Delta_{42} &\leq \sum_{j=m}^{J-1} \sum_{l \in U_j} [C_u C_\lambda R_{jl\varepsilon} \mathbb{I}(B_{jl} \geq 0.25 \tau^2 R_{jl\varepsilon}) + B_{jl} \mathbb{I}(B_{jl} \leq 1.25 \tau^2 R_{jl\varepsilon})] \\
&\leq C \sum_{j=m}^{J-1} \sum_{l \in U_j} \min(B_{jl}, R_{jl\varepsilon}) \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho
\end{aligned} \tag{10.36}$$

by Lemma 7. Combination of formulae (10.28) – (10.36) completes the proof of (5.4).

In order to prove (5.6), note that

$$\Delta^* = \mathbb{E} \|\hat{f}_{c,m} - f_c\|^4 \leq \Delta_1^* + \Delta_2^* + \Delta_3^*,$$

where

$$\begin{aligned}\Delta_1^* &= O\left(\mathbb{E}\left\|\sum_{k \in K_{0m}^c} (\hat{a}_{mk} - a_{mk}) \varphi_{mk}\right\|^4\right), \quad \Delta_2^* = O\left(\left\|\sum_{j=m}^{\infty} \sum_{k \in K_{1j}^c} b_{jk} \psi_{jk}\right\|^4\right), \\ \Delta_3^* &= O\left(\mathbb{E}\left[\sum_{j=m}^{J-1} \sum_{l \in U_j} \sum_{k \in U_{jl}} (\tilde{b}_{jk} - b_{jk})^2 \mathbb{I}(\hat{B}_{jl} \geq 0.25\tau^2 R_{jl\varepsilon})\right]^2\right).\end{aligned}$$

Observe that, by Assumption (A4), since $m \leq J$,

$$\begin{aligned}\Delta_1^* &= O\left(2^m \sum_{k \in K_{0m}^c} \mathbb{E}(\hat{a}_{mk} - a_{mk})^4\right) = O\left(2^m \varepsilon^2 \sum_{k \in K_{1j}^c} \lambda_{m,k}^{-4}\right) \\ &= O\left(\varepsilon^2 2^{J(2\alpha+2\beta+1+(1-2\alpha)_+)} [\ln(\varepsilon^{-1})]^{\mathbb{I}(2\alpha=1)}\right) = O\left(\varepsilon^{-\frac{2\alpha+2\beta-2+(1-2\alpha)_+}{2+\alpha+\beta}}\right) = o(\varepsilon^{-2}).\end{aligned}$$

For Δ_2^* , by (10.15), we have

$$\Delta_2^* = O\left(\left[\sum_{j=m}^{\infty} \sum_{k \in K_{1j}^c} b_{jk}^2\right]^2\right) = O\left(2^{-4ms'}\right) = o(1).$$

Finally, similarly to (10.32), partition Δ_3^* as $\Delta_3^* = \Delta_{31}^* + \Delta_{32}^*$ with Δ_{31}^* and Δ_{32}^* corresponding to $\mathbb{I}(\sum_{k \in U_{jl}} (\hat{b}_{jk} - b_{jk})^2 > 0.25\tau^2 R_{jl\varepsilon})$ and $\mathbb{I}(\sum_{k \in U_{jl}} b_{jk}^2 > 0.25\tau^2 R_{jl\varepsilon})$, respectively. For Δ_{31}^* , applying (4.15), and (10.33), obtain, as $\varepsilon \rightarrow 0$,

$$\begin{aligned}\Delta_{31}^* &= O\left(2^J \sum_{j=m}^{J-1} \sum_{l \in U_j} \mathbb{E}\left[\sum_{k \in U_{jl}} |\hat{b}_{jk} - b_{jk}|^4 \mathbb{I}\left(\sum_{k \in U_{jl}} (\hat{b}_{jk} - b_{jk})^2 > 0.25\tau^2 R_{jl\varepsilon}\right)\right]\right) \\ &= O\left(2^J \sum_{j=m}^{J-1} \sum_{k \in K_{1j}^c} \left[\mathbb{E}|\tilde{b}_{jk} - b_{jk}|^8\right]^{1/2} \left[\mathbb{P}\left(\sum_{k \in U_{jl}} (\hat{b}_{jk} - b_{jk})^2 > 0.25\tau^2 R_{jl\varepsilon}\right)\right]^{1/2}\right) \\ &= O\left(\varepsilon^2 \sum_{j=m}^{J-1} \sum_{k \in K_{1j}^c} \lambda_{j,k}^{-4} \varepsilon^{\frac{\chi^2}{2}}\right) = O\left(\varepsilon^{\frac{\chi^2}{2} - \frac{2\alpha+2\beta-2+(1-2\alpha)_+}{2+\alpha+\beta}}\right) = o(\varepsilon^{-2}),\end{aligned}$$

similarly to Δ_{31}^* . Finally, $\Delta_{32}^* = o(\varepsilon^{-2})$ by considerations similar to those in the case of Δ_{32} in (10.36).

Proof of Lemma 2 It is easy to see that $\Delta = \mathbb{E}\|\hat{f}_{0,m} - f_{0,m}\|^2 = \Delta_1 + \Delta_2$ where

$$\Delta_1 = \sum_{j=m}^{\infty} \sum_{k \in K_{0m}} b_{jk}^2, \quad \Delta_2 = \sum_{k \in K_{0m}} \mathbb{E}(\hat{a}_{mk} - a_{mk})^2, \quad (10.37)$$

and $\hat{a}_{mk} = \hat{z}_k^{(m)}$ for $k \in K_{0m}$. From characterization (3.1) of Besov spaces, it follows that, for any k , one has $b_{jk}^2 \leq A2^{-2js'}$, and, therefore, since the number of indices in the set K_{0m} is finite,

$$\Delta_1 = O\left(\sum_{j=m}^{\infty} 2^{-2js'}\right) = O\left(2^{-2ms'}\right). \quad (10.38)$$

In order to find an upper bound for Δ_2 , note that

$$\Delta_2 \leq 3(\Delta_{21} + \Delta_{22} + \Delta_{23}) \quad (10.39)$$

where

$$\begin{aligned} \Delta_{21} &= \mathbb{E} \|(\mathbf{A}^{(m)})^{-1}(\hat{\mathbf{c}}^{(m)} - \mathbf{c}^{(m)})\|^2, \\ \Delta_{22} &= \mathbb{E} \|(\mathbf{A}^{(m)})^{-1}\mathbf{B}^{(m)}(\hat{\mathbf{h}}^{(m)} - \mathbf{h}^{(m)})\|^2, \\ \Delta_{23} &= \|(\mathbf{A}^{(m)})^{-1}\mathbf{r}^{(m)}\|^2. \end{aligned} \quad (10.40)$$

Consider matrix $\mathbf{D}^{(m)} = \sqrt{\text{diag}(\mathbf{A}^{(m)})}$ with elements $\|w_{m,l}\|$, $l \in K_{0m}$, and matrix $\mathbf{G}^{(m)} = (\mathbf{D}^{(m)})^{-1}\mathbf{A}^{(m)}(\mathbf{D}^{(m)})^{-1}$. Note that $\mathbf{G}^{(m)}$ is a positive definite matrix of a finite (non-asymptotic) dimension with the unit main diagonal. hence, for some positive constants C_{G1} and C_{G2} independent of m , one has $\|\mathbf{G}^{(m)}\| \leq C_{G1}$ and $\|(\mathbf{G}^{(m)})^{-1}\| \leq C_{G2}$ which, in combination with (5.7), immediately imply (5.8). By (5.8), one has

$$\Delta_{21} \leq \|(\mathbf{A}^{(m)})^{-1}\| \mathbb{E} \|\hat{\mathbf{c}}^{(m)} - \mathbf{c}^{(m)}\|^2 \leq C \varepsilon \rho_m^{-4} \sum_{k \in K_{0m}} \|w_{m,k}\|^2 \leq C \varepsilon \rho_m^{-2}. \quad (10.41)$$

Now, let us examine Δ_{22} term. It is easy to see that

$$\Delta_{22} \leq C \|(\mathbf{D}^{(m)})^{-1}\| \|(\mathbf{G}^{(m)})^{-1}\| \mathbb{E} \|(\mathbf{D}^{(m)})^{-1}\mathbf{B}^{(m)}(\hat{\mathbf{h}}^{(m)} - \mathbf{h}^{(m)})\|^2.$$

It follows from Assumption (A4), formula (5.7) and condition (5.10) that

$$\begin{aligned} \Delta_{22} &\leq C \varepsilon \rho_m^{-2} \sum_{l \in K_{0m}} \sum_{k_1, k_2 \in K_{0m}^c} \|w_{m,l}\|^{-2} \langle w_{m,l}, w_{m,k_1} \rangle \langle w_{m,l}, w_{m,k_2} \rangle \langle t_{m,k_1}, t_{m,k_2} \rangle \\ &\leq C \varepsilon \rho_m^{-4} \sum_{l \in K_{0m}} \left\| \sum_{k \in K_{0m}^c} \langle w_{m,l}, w_{m,k} \rangle t_{m,k} \right\|^2 \leq C \varepsilon \rho_m^{-4} \sum_{l \in K_{0m}} \sum_{k \in K_{0m}^c} \lambda_{m,k}^{-2} \langle w_{m,l}, w_{m,k} \rangle^2 \leq C \varepsilon \lambda_m^{-2}. \end{aligned}$$

Hence, due to condition (5.9),

$$\Delta_{22} \leq C \varepsilon \rho_m^{-2}. \quad (10.42)$$

Finally, since, due to (10.15), $|b_{jk}| \leq A 2^{-js'}$, observe that, by (5.8), one has

$$\begin{aligned} \Delta_{23} &\leq C_{A2}^2 \rho_m^{-4} \sum_{l \in K_{0m}} \langle w_{m,l}, QR_m \rangle^2 \\ &\leq C_{A2}^2 A^2 2^{-2ms'} \rho_m^{-4} \sum_{l \in K_{0m}} \left[\sum_{j=m}^{\infty} \sum_{k=0}^{2^j-1} |\langle w_{m,l}, v_{j,k} \rangle| \right]^2 \leq 2DC_{A2}^2 A^2 K_2 2^{-2ms'} \end{aligned} \quad (10.43)$$

Combination of (10.38) – (10.43) completes the proof of (5.12).

Now, we need to show that (5.13) holds. For this purpose, note that $\mathbb{E} \|\hat{f}_{0,m} - f_{0,m}\|^4 \leq \Delta_1^2 + \Delta^*$ where Δ_1 is defined in (10.37) and $\Delta^* = \Delta_1^* + \Delta_2^* + \Delta_3^*$ with $\Delta_1^* = \mathbb{E} \|(\mathbf{A}^{(m)})^{-1}(\hat{\mathbf{c}}^{(m)} - \mathbf{c}^{(m)})\|^4$, $\Delta_2^* = \mathbb{E} \|(\mathbf{A}^{(m)})^{-1}\mathbf{B}^{(m)}(\hat{\mathbf{h}}^{(m)} - \mathbf{h}^{(m)})\|^4$ and $\Delta_3^* = \|(\mathbf{A}^{(m)})^{-1}\mathbf{r}^{(m)}\|^4 = \Delta_{23}^2$. Recall that $\Delta_3^* = 0(1)$ as $\varepsilon \rightarrow 0$. It is also easy to check that, similarly to terms Δ_{21} and Δ_{22} , as $\varepsilon \rightarrow 0$, one has $\Delta_1^* \leq C \varepsilon^2 \rho_m^{-4} \leq C \varepsilon^2 \lambda_m^{-4}$ and

$$\Delta_2^* \leq C \varepsilon^2 2^{2m} \left[\rho_m^{-4} \sum_{k \in K_{0m}^c} \lambda_{m,k}^{-2} \langle w_{m,l}, w_{m,k} \rangle^2 \right]^2 \leq C \varepsilon^2 2^{2m} \lambda_m^{-4}.$$

Now, validity of (5.13) follows from the fact that, by (4.15),

$$\varepsilon^2 2^{2m} \lambda_m^{-4} \leq C \varepsilon^2 2^{J(2+\alpha+\beta)} [\ln(\varepsilon^{-1})]^{\mathbb{I}(\alpha=1)} \leq C \varepsilon^{-2}.$$

10.4 Proofs of statements in Section 6

Proof of Lemma 3. Note that by definition of \hat{m} , whenever $\hat{m} > m_0$, there exists $j > m_0$ such that $\|(\hat{f}_{m_0} - \hat{f}_j)\mathbb{I}(\Omega_{m_0})\|^2 > \kappa^2 \varepsilon \ln(\varepsilon^{-1}) \lambda_j^{-2}$. Therefore,

$$\mathbb{P}(\hat{m} > m_0) \leq \sum_{j=m_0}^{J-1} \mathcal{P}_j \quad \text{with} \quad \mathcal{P}_j = \mathbb{P}\left(\|(\hat{f}_{m_0} - \hat{f}_j)\mathbb{I}(\Omega_{m_0})\|^2 > \kappa^2 \varepsilon \ln(\varepsilon^{-1}) \lambda_j^{-2}\right). \quad (10.44)$$

Observe that since

$$\begin{aligned} \|(\hat{f}_{m_0} - \hat{f}_j)\mathbb{I}(\Omega_{m_0})\| &\leq \|(\hat{f}_{0,j} - f_{0,j})\mathbb{I}(\Omega_{m_0})\| + \|(\hat{f}_{c,j} - f_{c,j})\mathbb{I}(\Omega_{m_0})\| \\ &\quad + \|(\hat{f}_{0,m_0} - f_{0,m_0})\mathbb{I}(\Omega_{m_0})\| + \|(\hat{f}_{c,m_0} - f_{c,m_0})\mathbb{I}(\Omega_{m_0})\|, \end{aligned}$$

one has the following upper bound for \mathcal{P}_j defined in (10.44):

$$\mathcal{P}_j \leq \mathcal{P}_{0,j,m_0} + \mathcal{P}_{0,j,j} + \mathcal{P}_{c,j,m_0} + \mathcal{P}_{c,j,j} \quad (10.45)$$

where, for any $m_0 \leq m \leq j$,

$$\begin{aligned} \mathcal{P}_{0,j,m} &= \mathbb{P}\left(\|(\hat{f}_{0,m} - f_{0,m})\mathbb{I}(\Omega_{m_0})\| > 0.25 \kappa \eta_{j\varepsilon}\right), \\ \mathcal{P}_{c,j,m} &= \mathbb{P}\left(\|(\hat{f}_{c,m} - f_{c,m})\mathbb{I}(\Omega_{m_0})\| > 0.25 \kappa \eta_{j\varepsilon}\right). \end{aligned}$$

Since $\text{supp}(f_{0,m}) \subseteq \Omega_m \in \Omega_{m_0}$ for $m \geq m_0$, one has

$$\begin{aligned} \|(\hat{f}_{0,m} - f_{0,m})\mathbb{I}(\Omega_{m_0})\|^2 &= \|(\hat{f}_{0,m} - f_{0,m})\mathbb{I}(\Omega_m)\|^2 = \|\hat{f}_{0,m} - f_{0,m}\|^2 \\ &\leq \|\hat{\mathbf{z}}^{(m)} - \mathbf{z}^{(m)}\|^2 + 2DA^2 2^{-2ms'}. \end{aligned} \quad (10.46)$$

Hence, applying (10.46) and Lemma 6 with $\nu = \kappa/8$, one derives

$$\mathcal{P}_{0,j,m} \leq \mathbb{P}\left(\|\hat{\mathbf{z}}^{(m)} - \mathbf{z}^{(m)}\| > 0.25 \kappa \eta_{j\varepsilon} - A\sqrt{2D}2^{-js'}\right) \quad (10.47)$$

$$\leq \mathbb{P}\left(\|\hat{\mathbf{z}}^{(m)} - \mathbf{z}^{(m)}\| > \kappa \eta_{j\varepsilon}/8\right) \leq C\varepsilon^{\frac{\kappa^2}{2^6 C\nu}}, \quad (10.48)$$

since, $\sqrt{2D} A 2^{-js'} < \kappa \eta_{j\varepsilon}/8$ for $m_0 \leq m \leq j$ if ε is small enough.

Now, let us consider the second term, $\mathcal{P}_{c,j,m}$. Denote

$$\begin{aligned} L_D &= \min(L_\varphi - D_0, L_\psi - D), \quad U_D = \max(U_\varphi + D_0, U_\psi + D), \\ C_D &= \max(|L_\varphi - D_0|, |L_\psi - D|, |U_\varphi + D_0|, |U_\psi + D|) \end{aligned} \quad (10.49)$$

and observe that $\text{supp}(\varphi_{mk})$ and Ω_{m_0} have non-empty intersection only if $k \in \tilde{K}_{m,m_0}$, where

$$\tilde{K}_{m,m_0} = \{k : |k - k_{0m}| > D_0, \quad 2^{m-m_0} L_D - U_\varphi < k - k_{0m} < 2^{m-m_0} U_D - L_\varphi\}.$$

Similarly, $\text{supp}(\psi_{jk})$ and Ω_{m_0} have non-empty intersection only if $k \in \tilde{K}_{j,m_0}$, where

$$\tilde{K}_{j,m_0} = \{k : |k - k_{0j}| > D, \quad 2^{j-m_0} L_D - U_\psi < k - k_{0j} < 2^{j-m_0} U_D - L_\psi\}.$$

Therefore, $l \in U_j^*$ where

$$U_j^* = \{l : l \in U_j, |l| \leq l_{j\varepsilon}^*\} \quad \text{with} \quad l_{j\varepsilon}^* = C_D [\ln(\varepsilon^{-1})]^{-1} 2^{j-m_0}, \quad (10.50)$$

and, hence, for $m \geq m_0$, one has

$$\begin{aligned} \|(\hat{f}_{c,m} - f_{c,m})\mathbb{I}(\Omega_{m_0})\|^2 &\leq \sum_{j=m}^{\infty} \sum_{k \in \tilde{K}_{j,m_0}} b_{jk}^2 + \|\hat{\mathbf{h}}^{(m)} - \mathbf{h}^{(m)}\|^2 \\ &+ \sum_{j=m}^{J-1} \sum_{l \in U_j^*} \sum_{k \in U_{jl}} (\hat{b}_{jk} - b_{jk})^2 \mathbb{I}(\hat{B}_{jl} \geq \tau^2 R_{jl\varepsilon}). \end{aligned} \quad (10.51)$$

It follows from (10.15) that, if ε is small enough, one has

$$\sum_{j=m}^{\infty} \sum_{k \in \tilde{K}_{j,m_0}} b_{jk}^2 \leq A^2 (2C_D)^{(1-2/p)+} 2^{-2m_0 s'} \leq 2^{-5} \kappa^2 \eta_{m\varepsilon}^2. \quad (10.52)$$

Also, observe that

$$\mathbb{I}(\hat{B}_{jl} \geq \tau^2 R_{jl\varepsilon}) \leq \mathbb{I}\left(\sum_{k \in U_{jl}} (\hat{b}_{jk} - b_{jk})^2 > 0.25\tau^2 R_{jl\varepsilon}\right) + \mathbb{I}\left(\sum_{k \in U_{jl}} b_{jk}^2 > 0.25\tau^2 R_{jl\varepsilon}\right). \quad (10.53)$$

If ε is small enough, then, due to (10.52), one has $b_{jk}^2 \leq 0.25\tau^2 R_{jl\varepsilon}$, and, hence, the second indicator in (10.53) is the identical zero. Hence, it follows from formulae (10.51) – (10.53) that

$$\mathcal{P}_{c,j,m} \leq \mathbb{P}\left(\|\hat{\mathbf{h}}^{(m)} - \mathbf{h}^{(m)}\| > \frac{\kappa \eta_{j\varepsilon}}{4\sqrt{2}}\right) + \sum_{j'=m}^{J-1} \sum_{l \in U_{j'}^*} \mathbb{P}\left(\sum_{k \in U_{j'l}} |\hat{b}_{j'k} - b_{j'k}|^2 > \frac{\tau^2 R_{jl\varepsilon}}{4}\right). \quad (10.54)$$

Since components $\hat{h}_k^{(m)} - h_k^{(m)}$ of vector $\hat{\mathbf{h}}^{(m)} - \mathbf{h}^{(m)}$ are normally distributed with zero means and variances $\varepsilon \|t_{mk}\|^2$, using Assumption (A4), one obtains for the first term in (10.54):

$$\mathbb{P}\left(\|\hat{\mathbf{h}}^{(m)} - \mathbf{h}^{(m)}\| > \frac{\kappa \eta_{j\varepsilon}}{4\sqrt{2}}\right) \leq \sum_{k \in \tilde{K}_{m,m_0}} \mathbb{P}\left(|\hat{h}_k^{(m)} - h_k^{(m)}| > \frac{\kappa \sqrt{\varepsilon \ln(\varepsilon^{-1})} \|t_{mk}\|}{4\sqrt{2}C_t}\right) \leq C \varepsilon^{\frac{\kappa^2}{64C_t} - 1}.$$

Recalling that the sum in (10.55) has at most 2^J terms and applying (10.33), derive

$$\mathcal{P}_{c,j,m} \leq C \varepsilon^{\frac{\kappa^2}{64C_t} - 1} + C \varepsilon^{\chi^2 - \frac{2}{2+\alpha+\beta}}. \quad (10.55)$$

Combination of formulae (10.44), (10.45), (10.47) and (10.55) and definition (10.10) completes the proof.

Proof of Theorem 3. Observe that

$$\Delta = \mathbb{E}[\|\hat{f}_m - f\|^2] = \sum_{m=m_1}^{m_0} \mathbb{E}[\|\hat{f}_m - f\|^2 \mathbb{I}(\hat{m} = m)] + \mathbb{E}[\|\hat{f}_m - f\|^2 \mathbb{I}(\hat{m} > m_0)] \equiv \Delta_1 + \Delta_2$$

and consider terms Δ_1 and Δ_2 separately. Note that for any $m \in [m_1, m_0]$ one has

$$\mathbb{E}\|\hat{f}_m - f\|^2 \leq 2\mathbb{E}\|\hat{f}_{m_0} - f\|^2 + 2\mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x \in \Omega_m)\|^2 + 2\mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x \in \Omega_m^c)\|^2$$

where set Ω_m is defined in (6.1). By Theorem 2, obtain

$$\mathbb{E}\|\hat{f}_{m_0} - f\|^2 \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho$$

where $\Delta(\varepsilon)$ and ρ are defined in (3.3) and (5.5), respectively. If $\hat{m} = m \leq m_0$, then by definition of \hat{m} , since $2s'/(2s' + \alpha + \beta) \geq 2s/(2s + \beta + 1)$ for $\alpha \geq 1$, derive that

$$\mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x \in \Omega_m)\|^2 \leq \kappa^2 \varepsilon \ln(\varepsilon^{-1}) \lambda_{m_0}^{-2} \leq C \varepsilon^{\frac{2s'}{2s' + \alpha + \beta}} [\ln(\varepsilon^{-1})]^{1 + \mathbb{I}(\alpha=1)} \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^{1 + \mathbb{I}(\alpha=1)}.$$

Now, recall that Ω_m is defined in such a way that $\text{supp}(f_{0,m}) \in \Omega_m$ for any m and $\Omega_{j_1} \subset \Omega_{j_2}$ for $j_1 > j_2$, so that, for $m \leq m_0$, one has

$$\begin{aligned} \mathbb{E}\|(\hat{f}_m - f)\mathbb{I}(x \in \Omega_m^c)\|^2 &= \mathbb{E}\|(\hat{f}_{0,m} + \hat{f}_{c,m} - f_{0,m} - f_{c,m})\mathbb{I}(x \in \Omega_m^c)\|^2 \\ &= \mathbb{E}\|(\hat{f}_{c,m} - f_{c,m})\mathbb{I}(x \in \Omega_m^c)\|^2 \leq \mathbb{E}\|\hat{f}_{c,m} - f_{c,m}\|^2 \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho \end{aligned}$$

as $\varepsilon \rightarrow 0$. Observing that

$$\mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x \in \Omega_m^c)\|^2 \leq 2 \left[\mathbb{E}\|(\hat{f}_m - f)\mathbb{I}(x \in \Omega_m^c)\|^2 + \mathbb{E}\|(\hat{f}_{m_0} - f)\mathbb{I}(x \in \Omega_m^c)\|^2 \right],$$

combining all formulae above and noting that ρ in (5.5) is such that $\rho \leq 1 + \mathbb{I}(\alpha = 1)$, obtain that $\Delta_1 \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^{1 + \mathbb{I}(\alpha=1)}$ as $\varepsilon \rightarrow 0$.

By Lemmas 1 and 2, as $\varepsilon \rightarrow 0$, one has $\mathbb{E}\|\hat{f}_{0,m} - f_{0,m}\|^4 = o(\varepsilon^{-2})$ and $\mathbb{E}\|\hat{f}_{c,m} - f_{c,m}\|^4 = o(\varepsilon^{-2})$. Then, since, due to (6.7), one has $d > 2$ in (6.4) and (6.5), Lemma 3 yields

$$\Delta_2 \leq \sqrt{\mathbb{E}\|\hat{f}_m - f\|^4} \sqrt{\mathbb{P}(\hat{m} = m > m_0)} = O(\varepsilon^{d^2/2-1}) = O(\varepsilon), \quad (\varepsilon \rightarrow 0)$$

which completes the proof of Theorem 3.

10.5 Proofs of statements in Section 7

Proof of Proposition 1.

Note that $v_{j,k}(x) = \mu(x)V_{j,k}(x)$ where $v_{j,k}(x) = \int_0^1 q(x-t)\psi_{jk}(t)dt$. Denote by Ψ_l the l -th anti-derivative of ψ , i.e.

$$\Psi_l(t) = [(l-1)!]^{-1} \int_{L_\psi}^t (t-z)^{l-1} \psi(z) dz, \quad \Psi_l^{(l)}(t) = \psi(t). \quad (10.56)$$

Observe that for $0 \leq l \leq r$ one has $\Psi_l(L_\psi) = \Psi_l(U_\psi) = 0$ and also $\text{supp}(\Psi_l) = (L_\psi, U_\psi)$, hence, integrating $V_{j,k}$ by parts $(r-1)$ times, we derive

$$V_{j,k}(x) = (-1)^{r-1} 2^{-j(r-1/2)} \int_{L_\psi}^{U_\psi} q^{(r-1)}(2^{-j}x - k - z) d\Psi_r(t).$$

According to assumptions on q , the derivative $q^{(r-1)}(x)$ has one or several jump discontinuities at points $\tilde{x}_1, \dots, \tilde{x}_L$. Without loss of generality, we consider the case when $L = 1$ (there is only one jump discontinuity) and $\tilde{x}_1 = 0$ (due to periodicity, one can always achieve this by an appropriate shift). Hence, by integrating $V_{j,k}$ by parts one more time, we obtain

$$V_{j,k}(x) = (-1)^r 2^{-jr} \left[\int_0^1 q^{(r)}(x-t) 2^{j/2} \Psi_r(2^j t - k) dt - 2^{j/2} \Psi_r(2^j x - k) \Delta q^{(r-1)}(0) \right] \quad (10.57)$$

where $\Delta q^{(r-1)}(0) = q^{(r-1)}(0+) - q^{(r-1)}(0-)$ is the size of the jump of $q^{(r-1)}(z)$ at $z = 0$.

Recall that $x = 0$ is the only jump discontinuity of $q^{(r-1)}(x)$ and that

$$\Theta_{jk} = \text{supp} [\Psi_r(2^j x - k)] = (2^{-j}(L_\psi + k), 2^{-j}(U_\psi + k)).$$

Therefore,

$$V_{j,k} = V_{j,k,1} + V_{j,k,2} + V_{j,k,3}, \quad (10.58)$$

where

$$\begin{aligned} V_{j,k,1}(x) &= (-1)^r 2^{-jr} \mathbb{I}(x \in \Theta_{jk}^c) \int_0^1 q^{(r)}(x-t) 2^{j/2} \Psi_r(2^j t - k) dt, \\ V_{j,k,2}(x) &= (-1)^r 2^{-jr} \mathbb{I}(x \in \Theta_{jk}) \int_0^1 q^{(r)}(x-t) 2^{j/2} \Psi_r(2^j t - k) dt, \\ V_{j,k,3}(x) &= (-1)^{r+1} 2^{-jr} \mathbb{I}(x \in \Theta_{jk}) \Delta q^{(r-1)}(0) \sigma^{-1}(x) 2^{j/2} \Psi_r(2^j x - k). \end{aligned}$$

Then, $\|v_{j,k}\|^2 = M_1 + M_2 + M_3 + 2M_{23}$ where

$$M_i = \int_0^1 [V_{j,k,i}(x)]^2 \mu^2(x) dx, \quad i = 1, 2, 3, \quad M_{23} = \int_0^1 V_{j,k,2}(x) V_{j,k,3}(x) \mu(x) dx.$$

Since, for $x \in \Theta_{jk}^c$, $\Psi_{r_1}(2^j x - k) = 0$ and $q(x)$ is r_1 times continuously differentiable, one can use formula (10.57) for $V_{j,k}$ with r_1 instead of r , so that

$$V_{j,k,1}(x) = (-1)^{r_1} 2^{-jr_1} \mathbb{I}(x \in \Theta_{jk}^c) \int_0^1 q^{(r_1)}(x-t) 2^{j/2} \Psi_{r_1}(2^j t - k) dt. \quad (10.59)$$

Therefore, we derive

$$\begin{aligned} M_1 &\leq 2^{-j(2r_1+1)} \|q^{(r_1)}\|_\infty^2 \|\Psi_{r_1}\|_{L^1}^2, \\ M_2 &\asymp \mu^2(2^{-j}k) \int_{\Theta_{jk}} V_{j,k,1}^2(x) dx \leq C 2^{-j(2r+\alpha+1)} [|k - k_{0j}|^\alpha + 1] \|q^{(r)}\|_\infty^2 \|\Psi_r\|_{L^1}^2, \\ M_3 &\asymp \mu^2(2^{-j}k) \int_{\Theta_{jk}} V_{j,k,2}(x)^2 dx \asymp 2^{-j(2r+\alpha)} \int_{\Theta_{jk}} |z + k - k_{0j}|^\alpha \Psi_r^2(z) dz \asymp \frac{|k - k_{0j}|^\alpha + 1}{2^{j(2r+\alpha)}}, \end{aligned}$$

and $|M_{23}| \leq \sqrt{M_2 M_3} \leq C 2^{-j(2r+\alpha+1/2)}$. Hence, due to condition $2r_1 + 1 > 2r + \alpha$, the value of $\lambda_{j,k}$ is given by expression (7.10).

Validity of Assumption (A2) then follows from formula (7.9) and the fact that functions $u_{j,k}$ have bounded supports. Assumption (A4) can be verified in a similar manner. In order to show that Assumption (A3) holds, use decomposition (10.58) as above and note that it is sufficient to verify Assumption (A3) for each of the three functions $V_{j,k,1}$, $V_{j,k,2}$ and $V_{j,k,3}$. For $V_{j,k,2}$ and $V_{j,k,3}$, Assumption (A3) is satisfied since both functions have bounded supports. In the case of $V_{j,k,1}$, due to (10.59), one has $\lambda_{j,k}^{-2} \|V_{j,k,1}\|^2 \leq C 2^{-j(2r_1+1-2r-\alpha)} [|k - k_{0j}|^\alpha + 1]^{-1}$, so that

$$\sum_{k=0}^{2^j-1} \lambda_{j,k}^{-2} \|V_{j,k,1}\|^2 \leq C 2^{-j(2r_1+1-2r-\alpha)} 2^{j(1-\alpha)} \leq C 2^{-2j(r_1-r)}.$$

Therefore, Assumption (A3) holds by Cauchy inequality. Assumption (A5) is valid due to condition (E2).

For completion of the proof, it remains to check conditions (5.9)–(5.11) of Lemma 2. To start

with, we need to evaluate ρ_m in (5.9). For this purpose, note that under condition (7.4), one has

$$\begin{aligned}
\|w_{jk}\|^2 &= \int_0^1 \mu^2(x) dx \left[\int_0^1 2^{m/2} \varphi(2^m z - k) q(x - z) dz \right]^2 \\
&\asymp 2^{-m} \int_0^1 |y|^\alpha \left[\int_{L_\varphi}^{U_\varphi} \varphi(z) q(y - 2^{-m}(k - k_{0m} - z)) dz \right]^2 dy \\
&= 2^{-m} \int_0^1 |y|^\alpha \left[\int_{L_\varphi}^{U_\varphi} \varphi(z) \{q(y) - 2^{-m}(k - k_{0m} - z)q'(y - 2^{-m}[k - k_{0m} - \xi(z)])\} dz \right]^2 dy \\
&= 2^{-m} \int_0^1 |y|^\alpha q(y) dy \left[\int_{L_\varphi}^{U_\varphi} \varphi(z) dz \right]^2 + 2^{-m} R_m.
\end{aligned}$$

Here $R_m \leq C[2^{-m}|k - k_{0m}| \|q\|_\infty \|q'\|_\infty + 2^{-2m}|k - k_{0m}|^2 \|q'\|_\infty^2]$, so that, R_m is always bounded and, $R_m \leq C2^{-m}$ if $k \in K_{0m}$. Therefore, $\rho_m^2 = 2^{-m}$. It also follows from the above that quantities $\rho_m^{-4} \langle w_{m,l}, w_{m,k} \rangle^2$ are uniformly bounded above and, thus, condition (5.10) holds due to definition (5.1) and condition (5.9).

Now, it remains to verify condition (5.11). Note that $V_{j,k}$ in (10.57) can be expressed as $V_{j,k}(x) = 2^{j/2} H_j(2^j x - k)$ where

$$H_j(x) = \begin{cases} (-1)^r 2^{-jr} \left[\int_{L_\psi}^{U_\psi} q^{(r)}(2^{-j}(x - t)) \Psi_r(t) dt - \Psi_r(x) \Delta q^{(r)}(0) \right], & \text{if } x \in [L_\psi, U_\psi], \\ (-1)^{r_1} 2^{-jr_1} \int_{L_\psi}^{U_\psi} q^{(r_1)}(2^{-j}(x - t)) \Psi_{r_1}(t) dt, & \text{if } x \in [L_\psi, U_\psi]^c, \end{cases}$$

and $[L_\psi, U_\psi]^c = [0, 1] \setminus [L_\psi, U_\psi]$. Then,

$$|\langle w_{m,l}, v_{j,k} \rangle| \leq C 2^{-j(\alpha+1/2)} 2^{-m/2} \|q\|_\infty \|\varphi\|_{L_1} \int |y + k - k_{0j}|^\alpha |H_j(y)| dy$$

where $W_{jk} = \int |y + k - k_{0j}|^\alpha |H_j(y)| dy = W_{jk1} + W_{jk2} + W_{jk3}$. Here,

$$W_{jk1} = \int_{[L_\psi, U_\psi]^c} |y + k - k_{0j}|^\alpha |H_j(y)| dy \leq C 2^{j(\alpha-r_1)} \int_0^1 |q^{(r_1)}(z)| dz \int_{L_\psi}^{U_\psi} |\Psi_{r_1}(t)| dt \leq C 2^{j(\alpha-r_1)},$$

$$W_{jk2} = 2^{-jr} \int_{L_\psi}^{U_\psi} |y + k - k_{0j}|^\alpha \left| \int_{L_\psi}^{U_\psi} q^{(r)}(2^{-j}(x - t)) \Psi_r(t) dt \right| dy \leq C 2^{-j(r-\alpha)},$$

$$W_{jk3} = 2^{-jr} |\Delta q^{(r)}(0)| \int_{L_\psi}^{U_\psi} |y + k - k_{0j}|^\alpha |\Psi_r(y)| dy \leq C 2^{-j(r-\alpha)}.$$

Thus, $|\langle w_{m,l}, v_{j,k} \rangle| \leq C 2^{-jr} 2^{-(j+m)/2}$ and

$$\rho_m^{-2} \sum_{j=m}^{\infty} \sum_{k=0}^{2^j-1} |\langle w_{m,l}, v_{j,k} \rangle| \leq 2^m \sum_{j=m}^{\infty} 2^{-jr} 2^{-(j+m)/2} 2^j \leq C 2^{m(r-1)},$$

so that, due to $r \geq 1$, condition (5.11) is satisfied, which completes the proof.

References

- [1] Abramovich, F. and Silverman, B.W. (1998) Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, **85** 115–129.

- [2] Antoniadis, A., Pensky, M., Sapatinas, T. (2012) Nonparametric Regression Estimation Based on Spatially Inhomogeneous Data: Minimax Global Convergence Rates and Adaptivity. *ArXiv* 1106.4461v4.
- [3] Bissantz, N., Hohage, T., Munk, A., and Ruymgaart, F. (2007) Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numer. Anal.*, **45**, 2610–2636.
- [4] Bunea, F., Tsybakov, A. and Wegkamp, M.H. (2007). Aggregation for Gaussian regression. *Ann. Statist.* **35**, 1674–1697.
- [5] Cavalier, L., Golubev, Yu. (2006) Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. Stat.* , **34**, 1653-1677.
- [6] Cavalier, L., Golubev, Yu., Lepskii, O., Tsybakov, A. (2003) Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Theory Probab. Appl.*, **48**, 426-446.
- [7] Cavalier, L., Golubev, G.K., Picard, D., Tsybakov, A.B. (2002) Oracle inequalities for inverse problems. *Ann. Statist.*, **30**, 843-874.
- [8] Cohen, A., Hoffmann, M., Reis, M. (2004) Adaptive wavelet Galerkin methods for linear inverse problems. *SIAM Journ. Numer. Anal.*, **42**, 1479–1501.
- [9] De Silva, G. C., Yamasaki, T., Aizawa, K. (2009) Sketch-based spatial queries for retrieving human locomotion patterns from continuously archived GPS data. *IEEE Trans. on Multimedia*, **11**, 1240 - 1253.
- [10] Donoho, D.L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis* **2** 101–126.
- [11] Engl, H.W., Hanke, M. and Neubauer, A. (2000) *Regularization of Inverse Problems*. Kluwer Academic Publishers, Netherlands.
- [12] Gaïffas, S. (2005). Convergence rates for pointwise curve estimation with a degenerate design. *Mathematical Methods of Statistics*, **14**, 1–27.
- [13] Gaïffas, S. (2006). Sharp estimation in sup norm with random design. *Statistics and Probability Letters*, **77**, 782–794.
- [14] Gaïffas, S. (2007). On pointwise adaptive curve estimation based on inhomogeneous data. *ESAIM: Probability and Statistics*, **11**, 344–364.
- [15] Gaïffas, S. (2009). Uniform estimation of a signal based on inhomogeneous data. *Statistica Sinica*, **19**, 427–447.
- [16] Golubev, Y. (2010) On universal oracle inequalities related to high-dimensional linear models. *Ann. Stat.*, **38**, 2751–2780.
- [17] Gurdev, L.L., Dreischuh, T.N., and Stoyanov, D.V, (2002) High-range-resolution velocity-estimation techniques for coherent Doppler lidars with exponentially shaped laser pulses. *Applied Optics*, **41**, 1741–1749.
- [18] Harsdorf, S. and Reuter, R. (2000). Stable deconvolution of noisy lidar signals. In *Proceedings of EARSeL-SIG-Workshop LIDAR*, Dresden/FRG, June 16–17.

- [19] Hoffmann, M., Reiss, M. (2008) Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Stat.* , **36**, 310–336.
- [20] Johnstone, I.M., Kerkycharian, G., Picard, D. and Raimondo, M. (2004) Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society, Series B*, **66** 547–573 (with discussion, 627–657).
- [21] Johnstone, I.M. and Raimondo, M. (2004). Periodic boxcar deconvolution and Diophantine approximation. *Annals of Statistics* **32** 1781–1804.
- [22] Kalifa, J. and Mallat, S. (2003). Thresholding estimators for linear inverse problems and deconvolutions. *Annals of Statistics* **31** 58–109.
- [23] Kerkycharian, G., Lepski, O., and Picard, D. (2001) Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields*, **121**, 137–170.
- [24] Knapik, B.T., van der Vaart, A.W., van Zanten, J.H. (2011) Bayesian inverse problems with Gaussian priors. *Ann. Statist.*, **39**, 2626–2657.
- [25] Lepski, O.V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory of Probability and its Applications*, **35**, 454–466.
- [26] Lepski, O.V. (1991). Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimators. *Theory of Probability and its Applications*, **36**, 682–697.
- [27] Lepski, O.V., Mammen, E. and Spokoiny, V.G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors *Annals of Statistics*, **25**, 929–947.
- [28] Lepski, O. and Spokoiny, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Annals of Statistics*, **25**, 2512–2546.
- [29] Li, C.-T., Satta, R. (2011) On the Location-Dependent Quality of the Sensor Pattern Noise and Its Implication in Multimedia Forensics. *4th International Conference on Imaging for Crime Detection and Prevention (ICDP 2011)*, London, United Kingdom.
- [30] Mair, B., Ruymgaart, F. H. (1996) Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, **56**, 1424–1444.
- [31] Meyer, Y. (1992). *Wavelets and Operators*. Cambridge: Cambridge University Press.
- [32] Miller, F.P., Vandome, A.F., McBrewster, J. (2009) *Amplitude Modulation*. AlphaScript Publishing.
- [33] Pensky, M. and Sapatinas, T. (2009). Functional deconvolution in a periodic case: uniform case. *Annals of Statistics*, **37**, 73–104.
- [34] Pensky, M., and Vidakovic, B. (1999) Adaptive wavelet estimator for nonparametric density deconvolution. *Annals of Statistics*, **27**, 2033 – 2053.
- [35] Starck, J.L., and Pantin, E. (2002) Deconvolution in Astronomy: A Review. *Publ. Astronom. Soc. of the Pacific*, **114**, 1051–1069.
- [36] Tsybakov, A.B. (2009) *Introduction to Nonparametric Estimation*, Springer, New York.

- [37] van Schie, G., Karssemeijer, N. (2008) Detection of microcalcifications using a nonuniform noise model. In *Lecture Notes in computer Science, Digital Mammography: 9th International Workshop Proceedings, IWDM 2008 Tucson, Arisona*, Ed. Krupinski, E.A. **5116**, Springer, 378–384.
- [38] Weddell, S.J., Webb, R.Y. (2008) The Restoration of Extended Astronomical Images using the Spatially-Variant Point Spread Function. In proceeding of 23rd International Conference on Image and Vision Computing New Zealand, 26-28 November 2008. DOI:10.1109/IVCNZ.2008.4762100